

Critics of 'gene foods' report are avoiding the real issues

Have US critics of agricultural biotechnology chosen to attack the authors of a new study on the safety of genetically modified crops because they lack the arguments with which to counter the study's content?

The National Academy of Sciences had not even published its new report on the safety of genetically modified crops when opponents of transgenic technology — including one congressman, Democrat Dennis Kucinich of Ohio — were moving to try and ensure that its contents were discounted by the public and the news media.

The study provides a balanced and thoughtful contribution to the public debate that is unfolding in the United States on genetically modified foods (see page 693). While concluding that there is currently no evidence that the class of these foods that it considered — those that are genetically modified to protect them from pests — pose any special risk to the public, it outlines a number of steps the US government could take to improve their regulation.

However, the value that such a study might have in informing the general public about the scientific issues at stake was somewhat diluted by the energetic efforts of anti-genetics campaigners and congressman Kucinich to discredit the process that produced it. These efforts are best summarized in the words of Andrew Kimbrell, director of the Center for Food Safety, a pressure group in Washington: he dismissed the academy's work as "paid-for science".

The allegation that studies conducted by groups of scientists are corrupted by researchers' conflicts of interest is not a new one, least of all for the academy. Environmental groups have for years alleged that academy panels are rigged, since panel members are academics who invariably have financial ties with the government agencies and

industrial corporations that fund most of the work in their disciplines.

In the case of the latest study, however, the allegation seems threadbare. It is said that four of the eight practising academics on the study panel are in receipt of research funds from the agricultural biotechnology industry. But it is hard to see how this could be otherwise, unless the critics expect the academy to select scientists who have worked for years in agricultural biotechnology without getting funds from the industry.

Kimbrell also says that the National Academy of Sciences is clearly open to corporate influence because it accepts corporate donations to its endowment fund. But given the diversity of donors, the idea that any corporation influences National Research Council staff or panels by contributing to this fund is exceedingly far-fetched.

And the critics claim the study is flawed because its original staff director, Michael Phillips, left after six months to work for the Biotechnology Industry Organization. Phillips' career move probably made him a lot of money, but against the backdrop of Washington's many revolving doors, it hardly constitutes a major scandal.

This particular study was paid for, to the tune of about \$250,000, by the academy itself, in an effort to infuse some science into an important public debate. It was conducted under rules devised by the academy's members, comprising the United States' most distinguished scientists. Perhaps those who claim this process is corrupt, and call for its output to be rejected, do so because they lack the arguments or the ammunition to attack the study on its merits. ■

Framework lives, but does it learn?

As the European Commission starts to plan its next Framework research programme, flaws in the present one are emerging.

Basic science tests hypotheses, and applied science solves problems. A curious irony of the European Commission's multi-year Framework programmes for research is that they support applied research, but conduct themselves as a series of independent, hypothesis-driven social experiments.

The latest, fifth, Framework programme (FP5) tests whether scientists across Europe can emerge from their narrow scientific worlds and successfully focus their work on helping to solve some of Europe's major socio-economic problems.

From the outset, the prospects for the success of this approach did not look encouraging (see *Nature* 398, 1; 1999). Now, after little more than a year, they seem even less so; indeed, there is evidence that attempts to meet these new goals have merely increased the already daunting criteria that grant applications must meet (see page 695). The commission is not entirely to blame: it is, after all, but a servant of many political masters, each with their own interests to safeguard.

The various parties negotiated for years over FP5, and finally came up with a grand political vision intended to engage research in tackling many of the chronic European 'ills', from high unemployment to the

inflexibility of the workforce. The experiment that is FP5 has only three more years to run. But any lessons learnt may be only partially relevant to its successor, FP6, whose radically different governing hypotheses are already being formulated. Research commissioner Philippe Busquin, who took office last year, is preparing a discussion paper on FP6 to bring to the Council of Research Ministers in June.

Busquin's vision has an entirely different emphasis: it seeks to improve the integration of national and European Commission research efforts. The tools by which this could be achieved are not yet defined. But the sort of thing that may emerge is a move away from the funding of large numbers of relatively small, complex projects to a more coherent investment in critical mass, for example by supporting or networking centres of excellence.

Busquin's aim is for the commission to interweave its own research money with that of member states. This will be another experiment, and one whose conditions will inevitably be muddled by political negotiations. But at first glance the vision, as outlined by Busquin in January, has a reasonableness that was absent from FP5 from the start. ■





GM foods
Tuck in, says National Academy of Sciences advisory panel
p693



Space storm
Satellite network gave early warning of auroral wonders
p694



Patent battle
First blood to Oxford in scrap over microarray rights
p697



Celestial slip
Planet spotted in Taurus is probably an optical illusion
p699

Critics challenge Celera's claims over human genome sequence...

Washington

Celera Genomics last week reached a significant milestone in the race to sequence the human genome. At a US House of Representatives subcommittee hearing, the company announced that it has completed the raw sequencing stage of its project.

But experts from the rival publicly funded Human Genome Project (HGP) believe Celera may have stopped short of its original goal. They question the company's claims that it will assemble the human genome within six weeks, and raise doubts over whether Celera, by itself, will be able to annotate the sequence by the end of the year.

The controversy over Celera's latest claim centres on the 'shifting finish line' in the raw sequencing race. When the company first announced that it would take on the human genome in 1998, it said it would single-handedly sequence the genome 10 times over (10X coverage) — the repeats being necessary for higher accuracy. Celera revised this goal to 4X sequence coverage in January this year (see *Nature* 403, 119; 2000), saying that it would then combine its proprietary data with the HGP's public data.

But, according to sequencing experts such as Philip Green from the University of Washington in Seattle, the 'endpoint' that Celera announced last week seems closer to 3.3X coverage.

The company claims to have achieved 11X 'clone coverage', a measurement that is notably different from sequence coverage. In each unit of sequence coverage, machines read the entire length of every DNA fragment. But in clone coverage, sequencers only scan the ends of each fragment, skipping over the middle.

Clone coverage lends itself to Celera's strategy for genome sequencing. The company uses a 'whole-genome shotgun' approach, in which it blasts a complete piece of DNA into millions of fragments of varying sizes, reads only their ends, then uses computer algorithms to match up the overlaps.

According to the company, the paired clone ends allow the genome to be assembled "much more completely than single-stranded sequencing methods allow at comparable



All join hands? US science adviser Neil Lane (left) and Celera Genomics head Craig Venter agreed last week that public/private collaboration would be desirable in principle.

levels of sequence coverage". But critics such as Green argue that Celera needs a higher amount of clone coverage to achieve the level of sequence coverage that ensures completeness and accuracy.

The public project, by comparison, is

doing 'clone-by-clone' sequencing, reading each DNA fragment in its entirety, then placing it on a physical map.

The HGP's map serves as a scaffold on which both the public and the private projects can affix data. Robert Waterston, head of the

...as biotech debate splits along party lines

Washington

Last week's hearing on genomics, organized by the science, energy, and environment subcommittee of the US House of Representatives science committee (see above), was intended to educate committee members about progress in the field. But it ended up by politicizing the issue.

The hearing opened with a Republican denunciation of President Bill Clinton's recent remarks on gene patenting, whereas Democrats supported his remarks. It closed with a show of Republican support for the

biotechnology industry in general, and Celera Genomics in particular.

James Sensenbrenner, (Republican, Wisconsin), chair of the House Committee on Science, blasted Clinton for making remarks about access to raw sequence data. Clinton's comments upset the biotech stock market last month, as some initially interpreted them as an attack on gene patentability.

Sensenbrenner said public officials should be careful about issuing such statements. "These events highlight the need for increased sensitivity on the part of

government officials to the well-being of the high-tech sector," he said. But Jerry Costello (Democrat, Illinois) applauded the policy.

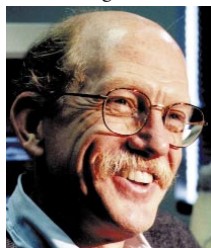
In closing, Ken Calvert (Republican, California) said that government's role is to initiate research on which private sector companies can capitalize, pointing to Celera Genomics as a prime example. He endorsed the way that Celera is incorporating the public Human Genome Project's data into its own database, adding its own sequence information, providing annotation tools, then selling it.

P. S.

department of genetics at Washington University in St Louis, told the committee that because Celera has access to that scaffold — as well as to the public project's data — it will always be ahead in the gene mapping effort.

Waterston believes that Celera's level of coverage, when assembled, will result in over 40,000 gaps. He based his estimate on Celera's sequence of the fruitfly *Drosophila*, which, at a fraction of the size of the human genome, contained about 1,200 gaps. Green says the number of gaps in Celera's human sequence might be higher, as the company did not sequence the genome as many times over as it did the *Drosophila* genome.

Nevertheless, Waterston, Celera's president Craig Venter and Neil Lane, President



Waterston: estimates 40,000 sequence gaps.

Bill Clinton's science adviser, all said at the subcommittee hearing that a formal collaboration would result in a better quality genome than either group could produce on its own. Indeed, each side may have something the other desires.

Although Celera has access to the public data, it has no guarantee that publicly funded scientists will volunteer to help annotate the human genome and fill in any gaps that remain once the two data sets are merged.

Public scientists participated in an 'annotation jamboree' for sequencing the genome of *Drosophila*. But that was on the understanding that the fruits of their labour would be available in GenBank, the publicly funded database. Celera has said it will make its version of the human genome available publicly, but only in its own database, and with restrictions on the use and redistribution of sequence data (see *Nature* 404, 324; 2000).

The public project, in turn, could use Celera's data to move up the completion date of its own project. The HGP has planned for a 'rough draft' of the genome with 5X sequence coverage by the end of this year, and a complete copy with 10X coverage in 2003.

Celera also plans to release its draft of the human genome by the end of this year — but only after it is fully assembled. By contrast, scientists can see new data being added to GenBank by the public project every 24 hours. This discrepancy between the two approaches has blocked formal cooperation from the beginning.

Celera's insistence that scientists should not be able to download, add to or redistribute information from either its prospective free or current public databases has placed further obstacles in the way of collaboration. Resolving them would allow for a more comprehensive human genome database sooner than either the public or private projects initially anticipated.

Paul Smaglik

German parliament agrees on limits to bioethics inquiry

Munich

The German parliament last week ended 18 months of controversy by deciding to set up an all-parliamentary *commission d'enquête* (commission of inquiry) on the ethical and legal aspects of biomedicine.

Opponents of such a commission had included Wolf-Michael Catenhusen, the biotechnology-friendly Social Democrat (SPD) secretary of research. He had feared that it would lead to cumbersome discussions of first principles that might delay legislation to bring the regulation of German biomedicine into line with the rest of Europe (see *Nature* 402, 331–332; 1999).

But the Greens — together with much of the media and the German public — had accused opponents of the commission of attempting to curtail political discussion about the ethical and social implications of biomedical and biotechnological progress.

Last week's compromise follows a reduction and streamlining of the inquiry panel's tasks. For example, the question of whether Germany should ratify the Council of Europe's Convention on Human Rights, which sets minimum ethical standards for European countries — but which Germany opposed because some clauses conflict with Germany's more restrictive national rules — has been removed from the agenda.

Catenhusen welcomes the compromise. "I am optimistic that the commission will focus on really important issues, such as

therapeutic cloning, now that we have turned away from setting up a combat group committed to fighting the Council of Europe Convention," he says.

The commission's purpose is now described as being to "work out recommendations for ethical assessment ... and for legislative and administrative action related to future medical opportunities". It will identify areas in which recent scientific developments have exposed a lack of appropriate legal rules, for example in stem-cell research or genetic testing.

Monika Knoche, the Green party's expert on medical ethics, says that genetic screening and reproductive technologies head the party's agenda for the commission.

The panel is scheduled to begin work next month, when its members — including 13 members of parliament and an equal number of external experts — will be nominated. It is not expected to include those holding 'extreme' positions. According to Catenhusen, this will reduce the chances of conflict found in the initial proposal, which would have included members known for their unwillingness to compromise.

Commissions d'enquête provide policy advice and prepare the introduction or modification of legislation. They are thus required to deliver final reports well before the end of a legislative period — a demand that, in the past, not all of them have been able to meet.

Quirin Schiermeier & Ulrike Hellerer

Ontario joins the genomics goldrush

Montreal

Canada's richest province, Ontario, is seeking to become an international force in genomics research. The Ontario Research and Development Challenge Fund (ORDCF) last week announced the approval of Can\$74.2 million (US\$51 million) for genomics-related projects. This sum will draw an additional Can\$134.7 million from research institutes and the private sector.

Ontario's plan predates the federal scheme to establish 'Genome Canada', comprising five national genome centres (see *Nature* 404, 8; 2000).

Last week's announcement of Ontario's projects was made by Jim Wilson, the province's energy, science and technology minister. Martin Godbout, named president of the national centre last month, has met with fund officials to discuss possible collaboration with Ontario's programme.

The projects approved for Ontario include a Centre for Genomic Computation, providing access to DNA sequences and analysis via the Internet. This will be housed at Toronto's Hospital for Sick Children, where new supercomputers have assumed management of the Genome Database of the international Human Genome Project. The database was transferred last year from Johns Hopkins University, Baltimore.

The Ontario Cancer Institute will host a proteomics facility, for producing pure proteins as drug targets, and a DNA microarray facility, for analysing information on the genetics of disease.

Each facility will serve the province as a whole. Seven more genomics proposals have been conditionally approved. Responses from the institutions involved, indicating whether all conditions have been met, are expected by the end of July.

David Spurgeon

US academy study finds GM foods are safe...

Washington

There is no scientific evidence that crops that have been genetically modified to resist pests pose special health or environmental risks, according to a study released last week by the US National Research Council (NRC), the research arm of the National Academy of Sciences.

But the study — commissioned and paid for by the NRC in a bid to inject scientific views into the debate in the United States on the safety of genetically modified (GM) food — still recommends that US government agencies tighten the regulation of GM food.

For example, it calls on the Environmental Protection Agency (EPA) to reconsider two broad exemptions excluding viral-coat proteins and genes from sexually related plants from the regulations that the agency imposes on GM pest-resistant plants.

"Although the committee believes that generally the system is working well, we have identified needed improvements," says Perry Adkisson, an entomologist and former chancellor of Texas A & M University, who chaired the NRC panel.

The study endorses the main conclusions of a 1987 National Academy of Sciences paper on the environmental impact of GM organisms (GMOs). This said that an organism's properties, rather than the process by which it was produced, were the key to its safety. This approach, which became key to the concept of 'substantial equivalence', formed the basis of the US government's regulation of GMOs.

However the new study calls for far more research into the environmental impacts of GMOs, and for greater coordination between the three US government agencies — the EPA, the Food and Drug Administration, and the Department of Agriculture — that regulate them. It calls on the agencies to establish a common database of GMO research.

Critics say that the report is inconsistent in endorsing substantial equivalence and then requesting special research into, and regulation of, GMOs. But panel members argue that public concerns make this necessary, even if GM foods are scientifically equivalent to traditional foods. "We need to make sure that the public has confidence in the technology," says Fred Gould, a panel member and professor of ecology at North Carolina State University.

Bruce Alberts, the president of the National Academy of Sciences, said last week that the study was the start of an effort to involve the scientific community in a broad public debate on GM foods. "It is very important that the voice of science be heard clearly," he said.

But even in advance of its publication, the study was dismissed by some environmental



Public outcry: biotech protesters hoping to make their voices heard at a rally in Boston in March.

groups, who claimed that the panel was biased (see below). Dennis Kucinich (Democrat, Ohio), a congressman who has been critical of agricultural biotechnology, said the panel "leans overwhelmingly toward a

pro-biotech position" and called on the academy to recall the report.

The biotechnology industry welcomed the findings. "This timely report... will reassure consumers on the thoroughness of the scientific scrutiny in place by US regulatory agencies," says Val Giddings, vice-president for food and agriculture at the Biotechnology Industry Organization.

NRC officials say the study is the first from the council to explicitly endorse the safety of GM foods. The 1987 study was concerned only with their impact on the natural environment.

The NRC also announced a new standing committee — co-chaired by Barbara Schaal of Washington University, St Louis, and Harold Varmus, former director of the National Institutes of Health and now president of the Memorial Sloan-Kettering Cancer Centre in New York — to "maintain surveillance" of agricultural biotechnology issues.

Colin Macilwain

...but critics claim the panel was biased

Washington

Even before the National Research Council (NRC) — the research arm of the National Academy of Sciences — released its report on genetically modified foods last week (see main story), pressure groups opposed to agricultural biotechnology were working to undermine it, arguing that panel members had conflicts of interest.

"The [academy] should hold the highest standards of independent scientific reporting, but this study absolutely does not meet those standards," said

Andrew Kimbrell, director of the Washington-based Center for Food Safety, which campaigns against agricultural biotechnology, in a statement distributed outside the academy's headquarters, where about 30 protesters gathered on the morning of the study's release. "The blatant conflicts of interest with the biotech industry put this study in the category of 'paid-for science'."

Environmental groups have been criticizing the study since last July, when Michael Phillips, the study's staff director, left to work for the Biotechnology Industry Organization. The NRC replaced him with another study



Kucinich: 'academy should scrap its GM food study'.

director, Jennifer Kuzma. Ever since, critics have called for the study to be abandoned, saying that the panel was biased.

In February, congressman Dennis Kucinich (Democrat, Ohio) wrote to Bill Colglazier, executive director of the academy, describing what he regards as conflicts of interest on the study panel, and asking for the study to be scrapped.

Kucinich and his supporters charge that six of the 12 panel members have received research funds from the biotechnology industry or done consultancy or legal work on its behalf. For example, they say that Fred Gould

has received research funds from Monsanto and Mycogen.

Gould points out that he has also taken money from the Union of Concerned Scientists and the Audubon Society, a conservation group. The literature distributed by groups opposing the study ignores this, say NRC officials, along with panel member Rebecca Goldberg's employment by the Environmental Defense Fund, a pressure group. The environmentalists say that Goldberg was added to the panel only after they objected to its original composition.

Eight of the panel are university researchers, and NRC says it is inevitable that they will have had grants from industry. Asked why he thought that the academy would rig its own panel, Kimbrell said it appeared to have been influenced by industrial contributions to its endowment fund.

"The academy allows itself to take outside money from corporations," he said. "Its panels are biased towards a corporate stance." An academy spokesperson denied that the contributions had had any impact on the composition of its panels, or their findings.

C. M.

Music software to come to genome aid?

Paris

Student use of a new 'killer' Internet application that allows anyone connected to the web to share music files stored on the hard disk of their own computer has become so heavy that US campuses want to ban the software for fear that it will saturate their academic Internet connections.

But some scientists are thinking of adopting the principles behind the so-called 'Napster' technology themselves. They believe these could herald a new era in distributed computing, and in particular solve the thorny problem of how the vast community of biologists can collaborate on assigning functions to the genes in the human genome.

Anyone connected to the Napster software, which can be downloaded from the Internet (<http://www.napster.com/>), can do a single search for a song across all the hard disks of other Napster users and download it directly from the user's computer.

When Lincoln Stein, a bioinformaticist at the Cold Spring Harbor Laboratory in New York, heard about Napster, he was struck by the parallels with his own work on writing software for a distributed sequence annotation system for the human genome (see <http://stein.cshl.org/das/>). Napster, he realized, can be used to find and distribute information located anywhere on the Internet.

Stein believes that annotation, which involves predicting which sequence stretches are genes and what their function might be, calls for a radically different global database structure from the centralized system used

for gene- and protein-sequence data. At present, users submit and retrieve records to a few central databases, such as GenBank.

But many scientists believe that annotation is too large a task for a few large genome centres. It is more of an art than a science, and up to half the predictions are wrong. No single group is likely to be able to produce a definitive version.

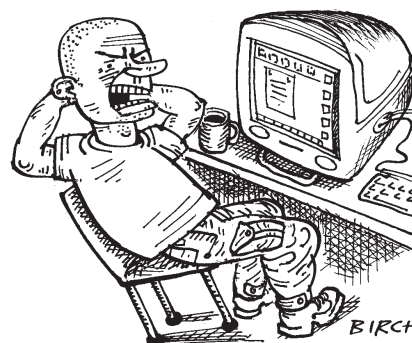
Annotation centres could in principle contribute data to GenBank-like centres. But GenBank entries, which can be modified only by those who submitted them, are sometimes erroneous. The problem is likely to be worse with the more subjective annotation data; the creation of a single authoritative annotated sequence seems unlikely.

A better solution, Stein argues, might be to allow biologists worldwide to annotate the human genome sequence interactively using diverse computational and experimental methods, much as developers worldwide debug open-source software.

But until now this sort of decentralized solution raised the spectre of duplication of effort, and the risk that scientists, instead of being able to consult a single central database, would have to search a series of separate versions of human genome databases. Data integration would become a real problem.

Stein believes Napster-like technology could be the answer. A centralized reference server holding a detailed genome map would act as an anchor for data produced locally by third-party annotation servers. Researchers could publish their data electronically with-

It's taking hours to download this new band, Human Genome Project...



out having to maintain their own websites.

Such a system could avoid the need to label entries with subjective identifiers, such as gene names or accession numbers, as occurs now. Instead, maps, gene predictions and functional activities could all be superimposed on the reference map using a system of coordinates, much as astronomers combine multiple-wavelength data with positional or coordinate information. A user could zoom in on any part of the genome and view the region in many ways by calling up related data from the hard disks of participating laboratories.

Ewan Birney, joint head of Ensembl, a joint venture between the European Bioinformatics Institute and the Sanger Centre in Cambridge to develop an automatic annotation on eukaryotic genomes, is backing Stein's idea, and has proposed that Ensembl be used as the reference map. The US National Center for Biotechnology Information, however, apparently opposes the idea on the grounds that it would lead to a proliferation of junk.

That risk is real. Although few genome scientists admit it publicly, the quality control of many smaller laboratories does not match that of specialized centres. "It's a Catch 22," says Birney. "We want to democratize people's ability to present their work, but quality is a problem." He believes one answer would be to offer users a full interactive choice between an approved annotation, composed of data from recognized gold-standard laboratories, and the full data.

Few researchers are willing to publicly endorse Napster-like technology. The idea of leaving desktop hard disks open to the Internet is a network manager's nightmare, and is not helped by Napster's ability to scan firewalls and breach weaknesses.

But Birney insists that developing the system should be taken seriously. "If we don't get it right it will be the difference between a biological web and people just continuing to use the Internet to send e-mails and read web pages."

Declan Butler

Space detectors show off their paces

Washington

Aside from lighting up night-time skies with colourful auroral displays last week, a magnetic storm that originated on the Sun acted as a showcase for the growing fleet of 'space weather' stations positioned at various points in the inner Solar System.

The first spacecraft to detect the storm was the Advanced Composition Explorer (ACE), parked in a stable orbit 1.5 million kilometres from the Sun. At 16:00 Universal Time (UT) on 6 April, ACE detected a shock wave associated with a gust in the solar wind of between 375 and 600 kilometres per second. The blast of atomic particles hit the Earth's magnetosphere less than an hour later and triggered the auroras.

Images from the US-European Solar and Heliospheric Observatory revealed the origins of the outburst in a solar flare that produced a large coronal-mass ejection at 15:41 UT on 4 April.



Light entertainment: the aurora as it was seen at the Brno Observatory in the Czech Republic.

According to a Space Weather Scale introduced last November by the US space agency NASA, this was a category 4 geomagnetic storm (category 5 is the highest). Storm activity is expected to pick up in the next couple of years as the Sun reaches its solar maximum.

Tony Reichhardt

Frustration grows over EU grant application procedures

Munich

Applying for research grants is always a complicated and time-consuming process. But many scientists across Europe are currently reserving their strongest complaints for applications to the European Union's (EU's) five-year Framework programme (FP5).

Not only are its application procedures especially cumbersome, claim researchers, but some say the chances of winning a grant are too low to justify the effort. Privately, some officials in the European Commission, which administers the programme, fear that the best researchers will be driven away.

Data emerging from the commission indicate that, after two rounds of proposals submitted last year, the success rate for applicants in the life sciences and information technologies — subjects earmarked by the EU as strategic priorities — remains below 20 per cent. In biotechnology, covered by a sub-programme called the 'Cell Factory', the success rate is as low as 10 per cent (see graph).

A Framework grant application involves recruiting a network of collaborating labs with an acceptable geographic balance: the chances of success are greater when researchers from smaller EU nations and from southern Europe are included. But FP5's requirement for projects to be justified in addition by their socio-economic value seems to be causing particular problems.

This requirement was agreed in 1998 by research ministers from the EU's member states and the European Parliament. Governments wanted FP5 to support research that directly addresses defined socio-economic

problems in European states, rather than simply supporting basic research, which, according to the EU's governing treaties, is the job of individual member states.

Scientists are struggling to put the idea into practice, however, and commission officials admit that the new criteria are difficult to explain, except by example. "You find yourself having to read through hundreds of pages of advice in order to construct an application," complains one frustrated, although successful, FP5 applicant.

Given the low success rates, many scientists are questioning the value of the time and effort spent on putting together an application. "For an application coordinator, an FP5 grant application is an extremely difficult and time-consuming affair that can take up to six months," says Hans-Georg Rammensee, a cell biologist at the University of Tübingen in Germany. "A rejection rate of some 90 per cent in my area makes this effort a huge waste of resources." Rammensee, although a partner in three Framework projects, says he would never volunteer to coordinate an application himself.

"You have to be desperate to apply for EU money," agrees Paula Ricciardi-Castagnoli, an immunologist at the University of Milan who has experienced the system as both a successful and an unsuccessful applicant and as a grant reviewer. "Scientists from countries with better funding opportunities like Germany and Britain will obviously turn to their national funding agencies where application procedures are less cumbersome."

A possible way around this problem would be to divert more FP5 money into life sciences and information technologies, in order to raise the success rate to the desired target of 30 to 40 per cent. But the EU's finance rules mean that, until FP5 expires at the end of 2002, money cannot be shifted from less popular research areas.

Instead, the commission is trying to tweak the system by cutting the number, and narrowing the scope, of future calls for proposals. It is also considering ways of making the application process more user-friendly.

Bruno Hansen, director of life sciences for FP5, says that simplification is a "high priority". But he defends the socio-economic criteria, arguing that this will guarantee that the European population gains maximum benefit from EU-funded research.

"I am very optimistic that it will prove fruitful, particularly in areas such as food safety or the development of new vaccines," says Hansen. "However, it is too early to say whether it is actually working."

Alison Abbott & Patrick Weydt



One in eleven: South Africa's government is reluctant to approve antiretroviral drugs.

Deaths bring South African HIV drug trials to a premature halt

Cape Town

South Africa's Medicines Control Council (MCC) has halted a clinical trial of the antiretroviral drug Coviracil (emtricitabine) following an announcement by the country's health minister that five South African women have died during the trial.

The decision has attracted particular attention because the study involved giving Coviracil to HIV-positive individuals in combination with d4T and nevirapine.

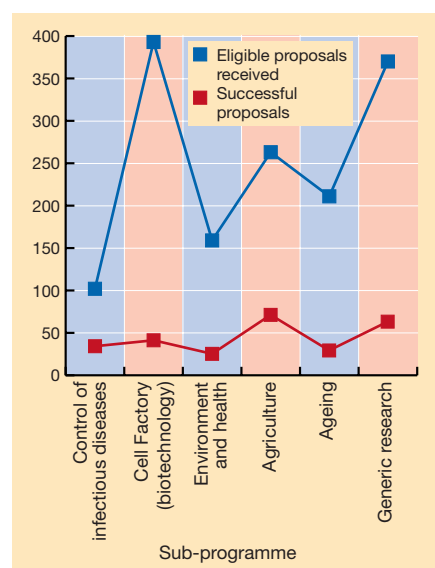
The government is already under pressure to allow HIV-positive pregnant women to be given nevirapine to prevent mother-to-child transmission of HIV, since President Thabo Mbeki has refused to sanction the use of AZT in state hospitals for this purpose (see *Nature* 402, 332; 1999).

Health minister Manto Tshabalala-Msimang has been accused of using the deaths to justify the government's refusal to provide pregnant women with antiretroviral drugs. "It is unfortunate that she has used the tragic event of deaths during the trials to make a political point that justifies her doing nothing to stop mother-to-child transmission," said opposition Member of Parliament Patricia de Lille of the Pan-African Congress.

But Tshabalala-Msimang told parliament last week that it would be "immoral and unethical" for the government to decide on the use of nevirapine until the full results of clinical trials on the drug are available.

MCC head Helen Rees denied that the government had put pressure on the council in making its decision. She described the timing of events as "serendipitous".

Michael Cherry



Slim pickings: EU grants require reams of paperwork, but the chance of success is small.

Director of Wellcome centre resigns over damning report

London

Roy Anderson, one of Britain's leading epidemiologists, has resigned as director of an Oxford-based research centre financed by the Wellcome Trust, following two damning reports on the way that the centre has been managed.

An independent management report, which criticizes both the University of Oxford and the trust for not keeping a close enough eye on the situation, will be a blow to both institutions.

The Wellcome Trust, which funds the Centre for the Epidemiology of Infectious Diseases at Oxford, this week released a management review and interim audit review of the centre. Both reviews began earlier this year when Anderson was suspended following a series of complaints about his behaviour during staff appointment procedures (see *Nature* 403, 353 & 695; 2000).

The reviews found that centre research staff supported by the trust had accepted commercial grants worth £250,000 from a number of sources — including the company IBHSC Ltd in which Anderson owns a third of the shares.

But Anderson had failed to make a full disclosure of his interests in his company IBHSC Ltd to either the trust or the university, as required of Wellcome grant holders.

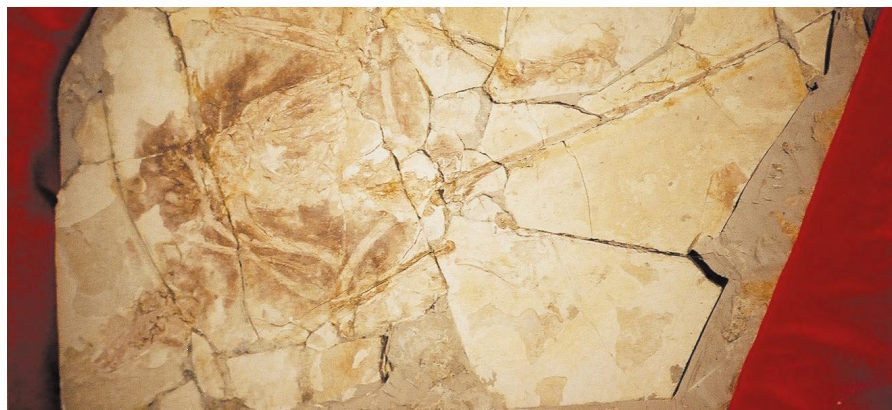
However, the interim audit, which was produced jointly by Price-waterhouseCoopers and the university's chief internal auditor, found no evidence of financial impropriety.

The independent management review was chaired by Sir Dai Rees, the former chief executive of the Medical Research Council. Its report says that the centre suffers from "autocratic management", festering tensions and disagreements, discouragement of independent views and growing tensions between the centre and the university's department of zoology to which it belongs.

Mike Dexter, director of the Wellcome Trust, says trustees were "obviously distressed" at the events surrounding the centre. He said there was "no question of the scientific credibility of Roy Anderson or the centre. It is doing first class work and our priority is to maintain that."

Anderson remains a governor of the Wellcome Trust, although Dexter says that his position "remains to be discussed".

Natasha Loder



Missing link: the 'Archaeoraptor' fossil is a composite, not a new species.

Fake bird fossil highlights the problem of illegal trading

Fort Lauderdale, Florida

It could have been an important clue to dinosaur and bird evolution. But a panel of palaeontologists last week confirmed that a fossil of a toothed bird — originally thought to be an important new species — is a composite of at least two separate specimens.

The fossil, which is thought to have been smuggled out of China, was bought early last year by amateur collectors, who paid \$80,000 for it at an Arizona mineral show (see *Nature* 403, 689; 2000). It will be returned to China in June as a result of negotiations completed last week while the fossil was being analysed at the US National Museum of Natural History in Washington DC.

According to several palaeontologists, the specimen still might have implications for the evolution of birds, and scientists at the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing are keen to study it.

But "it took about five minutes" to determine that the fossil was a composite, says Mark Norell, chairman of vertebrate palaeontology at the American Museum of Natural History in New York, and a member of the five-scientist panel.

According to Norell, the fossil tail of a primate bird had been inserted into the slab with the fossil remains of a more advanced bird, creating the appearance of a unknown specimen. Such fossils are typically altered in China to boost their value on the global underground market — they are classed as 'national treasures' in China, and cannot be sold legally.

The fossil — never named scientifically in a peer-reviewed journal, but known as 'Archaeoraptor' — was included in an article on bird evolution in *National Geographic* last November. The magazine had expected a paper on the specimen to appear in a scientific journal before its own article was published. But both *Nature* and *Science* declined to publish a submitted manuscript on the fossil.

National Geographic funded last week's analysis of the specimen, flying in a Chinese scientist and paying \$5,000 for the facing, counter-slab of parts of the 'Archaeoraptor' specimen. This counter-slab definitively proved the fossil's composite nature.

'Archaeoraptor' was discussed last week at the first Florida Symposium on Dinosaur Bird Evolution in Fort Lauderdale, sponsored by the Graves Museum of Archaeology and Natural History. But a lecture on the fossil by Stephen Czerkas, the amateur collector who first purchased it, threw little light on the controversy surrounding the specimen.

Indeed Czerkas, who directs a small museum in Blanding, Utah, also displayed and gave a talk on a second bird fossil. Several academic palaeontologists suggested this might also come from China, raising questions about whether the specimen had been smuggled out before Czerkas obtained it.

During two talks, Czerkas did not explain the origins of either specimen, or describe any scientific exchange with Chinese institutes. But he did publicly acknowledge for the first time that 'Archaeoraptor' was a composite.

Referring to the new fossil as "an arboreal theropod" with "remarkable implications", Czerkas claimed that it "represents a previously unknown lineage of dinosaurs that could climb".

But Kevin Padian, a curator at the University of California's Museum of Paleontology, said: "The idea that you look at a couple of features and say it lived in a tree is just not science. Such a determination requires a detailed study of all joints and motions." Czerkas refused to be interviewed after his talks.

The new specimen prompted Padian to call for aggressive new efforts by scientists and authorities to address the illicit trade in fossils. An international protocol is needed to keep the specimens in China "where they belong", he said.

Rex Dalton

Affymetrix loses first round of patent battle

London

An eminent Oxford professor has won the first round of a bitter battle with a California-based company that makes 'microarray' technology, used to determine which genes in a cell are switched on. Ed Southern claims that Affymetrix is trying to use its patents to exert a "stranglehold" over the development of the technology.

A British judge ruled last week that Affymetrix did not have the right to use techniques developed by Southern, perhaps best known as the inventor of the eponymous 'Southern blot' used for isolating and identifying DNA sequences.

Affymetrix had argued that it had acquired a license to basic microarray techniques patented by Southern in 1999, when it bought the microarray activities of Beckman Coulter, Inc. — which had entered into an earlier licensing agreement with the company Oxford Gene Technology (OGT), set up by Southern and the University of Oxford in 1995.

But the judge ruled that, despite its active research programme, Beckman Coulter did not have any products — and could not therefore be considered to be a 'business' in a

way that would have allowed Affymetrix to claim the rights to its licenses.

OGT officials say that their dispute with Affymetrix originated in Southern's commitment to licensing microarray technology in a way that gives competing research teams the freedom to develop rival techniques.

Southern says that he respects the work of Affymetrix's scientists. "But through their patents they are claiming rights over things that they did not invent, or that are not patentable."

The dispute over the transfer of the Beckman Coulter license follows the earlier breakdown of negotiations between Affymetrix and OGT. Chris Shelley of the Oxford company says that this was partly because of Affymetrix's refusal to allow OGT access to some of its own patents as part of a cross-licensing deal.

Last week's ruling will have a major impact on a broader case that is due to go to trial this October in a federal court in Delaware. In this, OGT is claiming that Affymetrix has infringed its patents in developing its so-called GeneChip technology. Affymetrix is responding by claiming that OGT's patents are invalid.



Southern: Affymetrix's scientists are "claiming rights over things that they did not invent".

The British company is also trying to get Affymetrix's British and European patents revoked. "The main basis of these actions is that Affymetrix's patents are unduly broad, covering areas of microarray technology that they cannot validly claim to have invented," say OGT officials.

Until now, Affymetrix's first line of defence against the charge of infringing OGT's patents was that its purchase of Beckman Coulter's 'business' gave it the right to use the OGT technology. The British judge's ruling, which is also valid in the United States, shoots down this argument.

OGT officials say that if their suit is successful, they will press for damages which — based on Affymetrix's sales figures — have been estimated at up to US\$300 million. Affymetrix officials last Friday defended their position, pointing out that the British ruling still left OGT having to defend its patents in the US and European courts.

"OGT will need to overcome significant flaws in their patents," says Vern Norviel, senior vice-president and general counsel of Affymetrix. "As the problems with OGT patents are exposed, we are highly confident that Affymetrix will prevail."

Norviel says that Affymetrix has "the strongest DNA patent portfolio in the field", with more than 70 patents issued and several hundred applications pending. "The ruling has no impact on Affymetrix's commanding intellectual property estate," he says.

Investors are hedging their bets. Some stock analysts are backing Affymetrix's confidence that it will be able to defend its strong patent position, but others are less optimistic. Last Friday, when most biotechnology stocks were rising, news of the British judgement sent Affymetrix's share price on the New York Stock Exchange down from \$143 to \$131.

David Dickson

UK ethicists back use of stem cells

London

British researchers are a step closer to working on human embryonic stem cells, following last week's endorsement of such research by the country's main bioethics panel, the Nuffield Council on Bioethics.

The endorsement coincides with press reports that an advisory panel set up by the British government last summer under its chief medical officer, Liam Donaldson, is also about to propose giving such experiments the green light.

Such a move would need parliamentary approval, as it would require the amendment of the regulations implementing the Human Fertilization and Embryo Act of 1990. These allow research on embryos up to 14 days old, but only for a defined set of purposes — which do not currently include stem-cell research.

The parliamentary debate would inevitably be stormy. Lord Alton, who as a Liberal Democrat Member of Parliament was among the fiercest critics of the 1990 act, wrote to *The Daily Telegraph* newspaper last week, describing the use of embryos to produce human 'spare parts' as "technological cannibalism".

But those in favour of a legislative amendment have generated a substantial

body of support. This now includes the Nuffield panel, which states that "there are no grounds for making a moral distinction between research into diagnostic methods or reproduction which is permitted under UK legislation and research into potential therapies which is not permitted".

Martin Bobrow, a member of the panel and professor of clinical genetics at the University of Cambridge, says that "we felt the potential of this research is so great as to warrant an extension of the categories of research that are permitted". However, he adds that, given the number of 'spare embryos' currently left over from fertility treatments, "we do not see the need for the creation of embryos for research purposes".

The BioIndustry Association, which represents the country's biotechnology companies, also supports a change in the regulation. John Sime, its chief executive, argues that the wide range of diseases that might be amenable to therapies based on embryonic stem-cell research means that "most families in Britain" could benefit.

Government officials argue that, even after the Donaldson panel's advice is made public, the decision on whether to amend the legislation should be left to the outcome of the parliamentary discussion.

D. D.

Record-breaking number of women doctorates awarded

Washington Women earned their largest percentage of doctorates in US history in 1998, according to a survey carried out for the National Science Foundation. Women received about 42 per cent of nearly 43,000 doctorates, including 45 per cent of those in the life sciences and 24 per cent in the physical sciences.

Over the past 40 years, the number of female doctorates has grown at more than twice the rate for men, at an average of 7.5 per cent a year. Among US citizens, ethnic minority groups earned 14.7 per cent of doctorates — also the largest percentage ever. The full report is available at <http://www.nsf.gov/sbe/srs/srs00410/htmstart.htm>

Security service guilty of theft from ecologist's flat

Moscow A local court in Vladivostok has ruled that officers of the Federal Security Service (FSB) broke the law when they took documents and personal possessions from the apartment of ecologist Vladimir Soyfer (see *Nature* 400, 300; 1999). Then head of the

nuclear oceanography laboratory at the Pacific Ocean Research Institute of the Far Eastern branch of the Russian Academy of Sciences, Soyfer was suspected by the FSB of revealing state secrets.

The chairman of the court had approved entry into Soyfer's apartment, but not a detailed inspection of the premises or the removal of his property. Photographs, an electronic notebook, computer disks, scientific papers and even personal correspondence were taken away, and Soyfer's passport was confiscated. An appeal by the local section of the security services against the court's ruling was rejected last week.

Nano-medicine gets the green light

Washington The US space agency NASA and the National Cancer Institute (NCI) are due to announce plans today (13 April) to collaborate on research into medical nanotechnology. The scheme will be aimed at developing nanometre-scale devices that can circulate in the bloodstream, diagnosing and possibly treating disease.

The NCI is interested in using the approach to diagnose cancer at an early stage; NASA believes that it could assist in monitoring the health of astronauts. The

initiative is expected to be announced by Dan Goldin, administrator of NASA, and Richard Klausner, director of the NCI, at a ceremony at the US Senate.

Question mark remains over French role in Diamond

London Uncertainty over French participation in Britain's new synchrotron source Diamond may continue until July. Although the French government announced last year that it would participate in the project, the recent resignation of science minister Claude Allègre — widely seen as the man responsible for the decision to participate in Diamond rather than constructing a French machine, Soleil — means that it remains possible, although unlikely, that France may pull out (see *Nature* 404, 533; 2000).

At a meeting last week between the British and French governments, it emerged that France will only be prepared to enter into a legal agreement once it has received a detailed costing for Diamond. Groups working towards such a legal agreement and detailed design specifications are expected to take several months to conclude their work. A union spokesperson for staff who will produce the design specifications says that such a study is not due to start until September.

'Planet' more likely to be stellar mirage

Washington What looked two years ago like the first planet photographed around another star is more probably an optical illusion, says the object's discoverer. Susan Terebey, an astronomer at Extrasolar Research Corporation in Pasadena, California, announced at a controversial press conference held by the US space agency NASA that the Hubble Space Telescope had photographed (see below) what appeared to be a planet flung away from

its parent star in the constellation Taurus (see *Nature* 393, 397, 406; 1998).

But follow-up spectroscopic observations with the Keck Telescope, along with temperature modelling of the object, known as TMR-1C, suggest that it is too hot to be a planet, and is more likely to be a dim background star. While the case is not closed, says Terebey, "at this time there is no strong evidence that TMR-1C itself is a protoplanet". Her conclusion will be published in the May issue of *Astronomical Journal*.

Foundations laid for radio telescopes of the future

Boston The SETI Institute and the Radio Astronomy Laboratory at the University of California, Berkeley, will next week inaugurate a prototype for the One Hectare Telescope (1hT), which, among other things, will be used in the search for extraterrestrial intelligence (SETI). The prototype, located at Berkeley's Leuschner Observatory, consists of a seven-telescope array that will begin operating on 19 April to test the hardware and software needed for 1hT.

"This is what we believe radio telescopes of the future will look like," says Tom Pierson, executive director of the SETI Institute. The 1hT will combine signals from 500 to 1,000 linked satellite dishes with a total collecting

area of 10,000 square metres. Construction of the telescope, which will be located at the Hat Creek Observatory in northern California, is due to start in 2003 and be completed in 2005.

Hennessy named as Stanford president

San Diego The new president of Stanford University is John Hennessy, a computer scientist, entrepreneur and the university's current provost. Hennessy, who taught at Stanford for 23 years, will succeed Gerhard Casper at the end of August.

Hennessy is perhaps best known for starting a project at Stanford in 1981 that developed the computer language RISC (Reduced Instruction Set Computer), which had a dramatic impact on the computer industry. In 1984, he co-founded MIPS Computer Systems, now known as MIPS Technologies, which makes microprocessors.

Correction

In our article 'Japan squares up to conservationists over grey whale's status' in last week's issue (see *Nature* 404, 531–532; 2000), the Japanese representative of the International Fund for Animal Welfare should have been named as Naoko Funahashi, and not Naoko Nakamae as printed.

S. TEREBEY/NASA



Mistaken identity: heralded as the first planet around another star, TMR-1C may be an illusion.

Storming the Tower of Babel

The search for life among the stars stepped up a gear last week with the First Astrobiology Science Conference in California. Henry Bortman and Philip Ball were there to test the atmosphere.

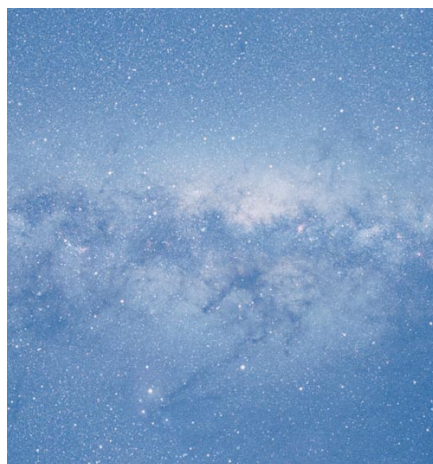
If scientific disciplines are usually conceived naturally, as like-minded researchers converge on an emerging idea, then astrobiology is more like a test-tube baby. This new field arose out of NASA administrator Dan Goldin's desire to make the search for extraterrestrial life one of the central themes for his agency.

Last week, some 500 scientists gathered at NASA's Ames Research Center in Moffett Field, California, to help map out an agenda for the newly emerging field. Most left cautiously optimistic about the discipline's future, but facing a difficult challenge: how to unite the diverse strands of research collected under astrobiology's umbrella. And, in particular, how to get the various specialists, each with their own research agendas and jargon, talking to one another in a meaningful way.

Breaking the mould

Some of the scientists at last week's meeting had gathered together before. But at earlier conferences, such as the six Bioastronomy meetings, they attended separate sessions where, for example, biologists would speak to biologists. The First Astrobiology Science Conference tried to break this mould with cross-disciplinary sessions along themes such as "Water: the *sine qua non* of life".

This made for a diverse set of presentations. Jody Deming, from the University of Washington in Seattle, discussed a new technique for studying the microorganisms that live inside tiny hypersaline pockets of brine within Arctic sea ice. These 'extremophiles' can survive at temperatures as low as -15°C , and are interesting because they live in conditions that might resemble those on Europa, one of Jupiter's icy moons.



Barren: is our galaxy's core devoid of life?

Bob Haberle of NASA Ames, meanwhile, presented a climate model for Mars. He showed that local pressure and temperature conditions on 29 per cent of the planet's surface could, in theory, enable liquid water to exist on the surface for minutes or maybe even hours at a time.

Other researchers considered the prospects for extraterrestrial life on a grander scale. While the 'habitable zone' of our Solar System is a well-established concept, Peter Ward, Donald Brownlee and Guillermo Gonzalez from the University of Washington have assessed the chances of such zones existing on a galactic scale. They argued that the outer reaches of a galaxy similar to our Milky Way would not contain sufficient quantities of heavy elements, such as silicon, aluminium and iron, to form the rocky, tectonically active planets thought to be a prerequisite for life. Near the centre of such galaxies, the dense

packing of stars means that planets are likely to be heavily bombarded by comets and bathed in hazardous radiation from supernovae. Most galaxies, the team concluded, will contain only limited regions where life is possible, and many may be utterly barren.

Teething troubles

Opinions differ on how successful the meeting was at integrating such diverse approaches into a coherent whole. "Differences in approaches between disciplines can provide fresh insights," says Jack Lisauer, a planetary scientist at NASA Ames. "If you look at a planet from a purely planetary-science perspective, you can look at 99.9 per cent of the mass, but can be missing the atmosphere completely. Yet that is crucial to habitability," he says.

On occasions, however, it was clear that communications had broken down. For instance, after one detailed presentation on evolutionary relationships among the Archaea, thought to represent some of the most primitive organisms on Earth, one bemused astronomer rose from his seat to ask: "Why should I be interested in this?"

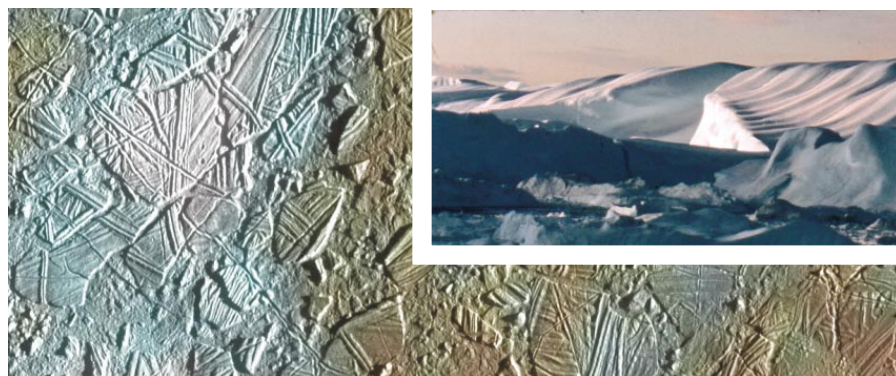
Given the circumstances of the field's genesis, some delegates at last week's meeting believe such glitches are inevitable. "Astrobiology is kind of an artificial conglomeration," says Jay Melosh, a theoretical geophysicist at the University of Arizona in Tucson. "Unlike biochemistry, which grew organically, this field was invented by Dan Goldin saying: 'This is going to be the field of astrobiology'."

Coordinating the field will be the task of NASA's Astrobiology Institute, an 'institute without walls' based at Ames but embracing a diverse set of research groups across the United States. While its task may be formidable, last week's meeting showed that astrobiology already has a respectable number of converts. Lynn Rothschild, a biologist at Ames who organized the meeting, says that attendance was twice as high as she originally anticipated.

"This may be remembered as the great ideas conference," concludes Melosh. "Out of that diversity there will be some ideas that get put together and perhaps grow and bear fruit. I'm not sure I can predict what's going to happen and what's not, but it will be interesting to watch."

Henry Bortman is a freelance writer in San Francisco; Philip Ball is a Consultant Editor of *Nature*.

♦ <http://astrobiology.arc.nasa.gov/conferences/2000/ABSciConf/index.html>



Life in the freezer: conditions in Arctic ice (inset) may mirror those on the surface of Europa.

Technology failures were caused by managers not listening to engineers

Sir—I do not agree with Sheila Jasanoff's point in her Millennium Essay "Knowledge elites and class war" about the dangers of technology¹. Of the three technological disasters mentioned by Jasanoff—Bhopal, Chernobyl and Challenger—the last two (and perhaps even the first) had a common root cause, in that the advice of expert engineers and risk analysts either was not sought or was overruled by management.

In the Challenger case, for example, the 'expert assessments' of the engineers were not incorporated into a proper risk analysis by NASA; engineers' recommendations were overruled by less expert management; and the astronauts were not made aware of the engineers' worries or of the disagreement. (See Richard Feynman's account².) NASA management procedures have much improved since then, but not to the extent of having an external expert safety review. In 1989, for example, an interagency safety review committee for the Galileo space probe was not informed about concerns expressed by a scientist at a lower level. This disagreement surfaced when the scientist wrote a letter directly to the president.

Chernobyl arose because the design assumed perfect operation. It was the compartmentalization of Soviet society and a politically based management that prevented information about design faults from reaching the operators.

It now seems clear that Bhopal arose from intentional sabotage by a disgruntled employee. Neither Union Carbide management nor the Indian government seems to have required a proper expert safety review of the plant by modern risk-assessment standards. Why, for example, was such an easily sabotaged plant designed and allowed to operate? Why were the operators not trained in the obvious remediation measures? This was a failure of management, not of the experts.

I do agree with Jasanoff that compartmentalizing society is dangerous, but that is becoming less prevalent. The crucial divisions are not those between workers and management or between expert and layperson, but those between management and experts—and between different experts, in societies in which management keeps them apart in order to control them.

The problem is compounded by politicians' refusal to recognize that the knowledge of scientific and technical experts must be at the heart of every scientific and technical decision they make.

There is no simple rule as to what defines an expert. But their role in judicial issues has recently been clarified in three landmark decisions: Daubert³, Joiner, and Kumho⁴. The first two of these were on admissibility of evidence of causation in medical situations, and the last on causation in engineering.

In Daubert the Supreme Court stated some non-exclusive criteria by which the court may judge the testimony being proffered. Has the theory been tested, or can it be tested? (In other words, is it falsifiable?) Has the theory been peer-reviewed and published? What is the known or potential risk of error? Has the theory been generally accepted in the relevant scientific community? Is the theory based on facts or data of a type reasonably relied upon by experts in the field? And, most importantly, does the testimony have probative value that is greater than, or not outweighed by, a danger of unfair prejudice, confusion of issues, or misleading the jury?

In Kumho the court declared that these criteria apply with greater force to situations in which an expert relies more on experience than on clearly defined theories.

DNA fingerprinting, despite being a relatively new technique, was not successfully challenged during the O. J. Simpson trial—though whether the jury accepted it is, of course, another matter.

It is important to incorporate expert knowledge into societal decisions. But a "conversation between science and society", as suggested by Jasanoff, is a peculiar recommendation. Science is an integral part of society, whether non-scientists like it or not. Some segments of society (including many non-scientists) understand the role science and technology play within society. That is the understanding that should be enhanced.

Richard Wilson

*Department of Physics, Harvard University,
Cambridge, Massachusetts 02138, USA*

1. Jasanoff, S. *Nature* **401**, 531 (1999).
2. Feynman, R. *What Do You Care What People Think? Further Adventures of a Curious Character* (Norton, New York, 1988).
3. Daubert v. Merrell Dow Pharmaceuticals (92-102), 509 U.S. 579 (1993); <http://supct.law.cornell.edu/supct/html/92-102.ZS.html>.
4. Kumho Tire Co vs Carmichael (97-1709). 131 F.3d 1433, reversed (1998–1999); <http://supct.law.cornell.edu/supct/html/97-1709.ZS.html>.

No conflict between SLAC and Japan's KEK

Sir—In the News profile on B-factories, my statements and those of others at Stanford Linear Accelerator Center (SLAC) were distorted (*Nature* **403**, 586; 2000). It is damaging to the relationship

between our laboratory and the Japanese KEK B-factory, and to me personally, to have my words made to appear competitive and inflammatory when that was not my intention.

Accelerators are complex machines that cannot be expected to operate perfectly as soon as they are turned on—reaching design performance is a process of gradual improvement.

I did say to your reporter that SLAC's machine worked very well in its first months and that everyone was surprised and pleased at how the performance had improved so quickly. I doubt that I used the word "debugging" with reference to this process.

I did not tell your reporter that there was any indication of whether the US or the Japanese experiment was ahead. I do not believe that it is possible to know this at the moment. Both projects have shown significant successes and both need much more work before the important physics results can be obtained.

One point not made in your article is the value of having two projects working on this same physics.

It is not true that only the first measurement will matter. In science, replicability of results is essential, and as a theoretical physicist I will be much more confident in the results if both experiments make similar findings.

On the same page, your reporter makes similar mistakes in commenting on the absence of the Fermilab director from the celebration honouring Burton Richter. The article implies this is because of bad blood; in fact it was simply because the Fermilab director was out of the country for personal matters that could not be rescheduled, a fact that could easily have been checked.

Helen Quinn

*Stanford Linear Accelerator Center, MS 81, PO Box
4349, Stanford, California 94309, USA*

Colin Macilwain replies—The opening paragraph of my story erred in stating that SLAC was "poised to win" its race with KEK to obtain significant measurements of charge-parity violation, but I stand by the rest of it, which accurately reflects the views of several scientists I interviewed at SLAC. The SLAC researchers believe themselves to be ahead at this early stage.

The director of KEK declined to be interviewed on this matter when I met him at SLAC on 21 January, but my colleague Robert Triendl obtained a response from Japan that confirmed my story's perspective, and was included in it.

Both of the quotes that I attributed to Dr Quinn in my story were accurate: her charges of distortion are not supported by the facts.

More than meets the eye

Chemistry, psychology, art and alchemy casting light on a multifaceted topic.

Color: A Multidisciplinary Approach

by Heinrich Zollinger

Wiley-VCH: 1999. 268 pp. £70, \$94.95

Philip Ball

Suppose some innovative university decided to offer a degree course in colour. They might, perhaps, argue that experts in colour are as useful to society as experts in ancient history, French literature, astrophysics or economics. And they might even be right. What, then, would the students be required to learn? Heinrich Zollinger's multidisciplinary book gives some inkling of the scope. Chemistry, certainly — although the emphasis it enjoys in this book no doubt reflects the author's own background. One would need a good grounding in physics, particularly optics and electromagnetism: Maxwell's equations, scattering theory, arguably even quantum electrodynamics. But then, perhaps colour is more about biology than physics — as Isaac Newton reminds us, light rays "are not coloured". It is, after all, the mechanisms of colour vision that create yellow from 'red' and 'green' light.

Then there is the psychology of colour, at least as useful to advertisers as to painters. Colour symbolism pervades art and literature, particularly in the Romantic tradition that Johann Goethe spawned and that Philipp Otto Runge, Arthur Schopenhauer, Rudolf Steiner and the Theosophists perpetuated in their various ways. Colour linguistics is a minefield, through which the hypothesis of basic colour terms posited by Brent Berlin and Paul Kay in 1969 (*Basic Color Terms: Their Universality and Evolution*, CSLI) offers a decidedly risky path. Ever since William Gladstone proclaimed, on the basis of the references to colour in classical literature, that the ancient Greeks were virtually colour-blind, we have learnt to beware of our own preconceptions in studying how other cultures think (and write) about colour.

Yet it is probably through art, not science, that the most inclusive study of colour should proceed. "I am a good half a scientist," said Henri Matisse, and one suspects the Neo-Impressionists Georges Seurat and Paul Signac would have claimed an even larger fraction. For Leonardo da Vinci, "Those who devote themselves to [artistic] practice without science are like sailors who put to sea without a rudder or compass and who can never be certain where they are going".

Says Zollinger, "it is a *conditio sine qua non* to include a chapter on color in art"; but is that enough? The history of the use of colour in art brings into focus our cultural attitudes



Japanese hues: *Moon Pine, Ueno*, a woodblock print by the nineteenth-century artist Hiroshige.

in all its respects. One can discern Aristotle's philosophy in the four-colour palettes of the most famous classical Greek painters, such as Apelles. Medieval altar paintings encode a symbolism that can be identified elsewhere in heraldry, alchemy and stained-glass work. Newton's influence shines out in John Constable's rainbows, and the halo in Matthias Grünewald's *The Resurrection* (c.1515) presages a physiological understanding of after-images in complementary colours. Paul Klee's *The Light and So Much Else* (1931) quantizes the light it portrays.

So I question whether Zollinger was right to root so much of his 'multidisciplinary' study in science — although the tradition for doing so includes Hazel Rossotti's eternally reprintable *Colour: Why the World Isn't Grey* (Princeton University Press, 1985). I suppose that, until universities do start teaching courses on colour, we are lumbered with our conventional predilections.

The real go-betweens are not, however, the physicists or the biologists, but the chemists — a notion with which I suspect Zollinger would gladly agree. As the abstract artist Wassily Kandinsky demonstrated, one can do astonishing things with colour while having only a poor grasp of its theoretical principles. (Indeed, as Kandinsky's one-time colleague Johannes Itten at the Bauhaus school of design showed, one can even lecture about it on that basis.) But without some knowledge of the chemical substance of

colour, the artist is highly vulnerable. Joshua Reynolds, Vincent van Gogh and Mark Rothko are among those who discovered this to the detriment of their works. The medieval painter, who typically made his own materials, was likely to be "a good half an alchemist" — if only to avoid being duped by unscrupulous apothecarists who might mix brick dust with vermilion or pass off azurite as precious ultramarine. It is no coincidence that one of the few sixteenth-century German painters to use the rare and toxic yellow pigment orpiment was Lucas Cranach, who ran his own pharmacy.

It may be more surprising to discover that this beneficial relationship between chemistry and art has been mutual. The modern chemicals industry was conceived and nurtured largely by the demand for colour. Advances in synthetic chemistry, both organic and inorganic, were stimulated in the nineteenth century by the quest for artificial colours. In Napoleonic France the chemist Louis-Jacques Thénard was directed by the government to create a synthetic substitute for expensive ultramarine, leading him to devise cobalt blue in 1802. William Perkin's mauveine in 1856 helped to launch both the 'Purple Decade' in visual art and design and, in the longer term, the entire synthetic dye-stuffs industry. Many of the world's major chemicals companies — BASF, Bayer, Hoescht, Ciba-Geigy — began their existence as manufacturers of aniline dyes related to those of Perkin. The serious commercial synthesis of natural products arguably began in 1868 with alizarin, the natural red dye from the madder plant that is used in the red pigment lake.

Zollinger covers an impressive amount of this territory. That he is an enthusiast and self-confessed autodidact in several areas has both its advantages — an infectious delight in the subject — and its disadvantages. Stylistically, the book is an odd mix, which the translation has not enhanced. Where the author's particular interests come into view, he is prone to descend into a textbook style or a research paper in disguise. The whole ride is, in fact, a jerky one, with puzzling detours, occasional congestion and crunching changes of gear, interspersed with spells of smooth cruising. And the decision to italicize all proper names is pursued with bizarre doggedness, so that even Queen *Victoria* and the *Weizmann* Institute get this treatment.

But the author's biggest hurdle is that he has chosen to write the book for the as-yet fictitious student of colour. Who else, chemists aside, will consider themselves obliged to wade through hard-core

book reviews

molecular-orbital theory before sampling the delights of Titian's *Bacchus and Ariadne* and Hiroshige's "One Hundred Famous Views of Edo"? Trimmed of such excesses, this attractive book might gain the popularity that its topic deserves. ■

Philip Ball is a consultant editor at Nature.

Endangered harvest?

Sustainable Use of Hawksbill Turtles: Contemporary Issues in Conservation

by N. Mrosovsky

Key Centre for Tropical Wildlife

Management: 2000. 107 pp. AU\$25 (pbk)

John G. Robinson and John Thorbjarnarson

For the current meeting of CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora; 10–20 April) in Nairobi, the Cuban government has submitted proposals to allow a small number of hawksbill turtle shells to be harvested from Cuban waters and exported to Japan. A CITES Appendix I listing has since 1976 banned the trade of this turtle's beautiful shell, which is prized for its use in ornaments and inlays, on the basis that the turtle populations were affected by international trade and that the species was threatened with extinction.

The Cuban proposals argued that hawksbills (*Eretmochelys imbricata*) can be sustainably harvested, and that active international trade would promote the turtle's conservation. These proposals have generated passionate controversy, and a balanced analysis is needed of whether such a strategy might be applied to hawksbills. Mrosovsky's book falls short of fulfilling this need.

The book, timed to influence the CITES debate, is an emotional plea in support of Cuba's position. It argues that there is a general lack of information on wild populations, so that effective conservation management is invariably experimental and adaptive. It also says that available information indicates that hawksbill turtles are not in danger of extinction, and that allowing the turtle trade, provided it is sustainable, is a better conservation strategy than strategies based on protection. Finally, the book states that institutional mechanisms are in place to manage the wild populations and the trade.

Mrosovsky's first point is well developed. Conservation decisions must often be taken without full knowledge of the situation, and sometimes the safest or most 'precautionary' step is bold experimentation. The book's other points, however, are less effectively argued. Mrosovsky's argument that certain

populations of hawksbills are showing recovery does not negate the consensus that the species population is still only a tiny fraction of its historical numbers. The fact that hawksbills do not fit the 'Critically Endangered' category of IUCN, the World Conservation Union, says nothing about their inclusion in CITES Appendix I. Mrosovsky's assertion that sustainable use works for crocodiles, elephants, whales and rhinos does not establish its applicability to hawksbills. His diatribes against protection as a conservation strategy ignore the fact that most recovering populations of hawksbills are those that are being effectively protected. While the Cuban authorities have a well-managed programme, this does not obviate the need for the regional management of this highly migratory species, nor for effective regulation of the trade in Japan.

A fundamental requirement of any sustainable-use programme is that such use promotes conservation. Revenues from the harvest might be used to manage the wild population; local fishing communities might adopt the role of stewards of the resource; and other competing industries might be excluded. Mrosovsky fails to make any of these arguments, and puts forward as the primary benefits of trade only the avoidance of waste from incidental catching of hawksbills, and the use of revenues to construct an education centre in Cuba. These reasons are not compelling.

Argument by analogy and innuendo, disregard for scientific credentials, questioning the pecuniary motives of real and imaginary opponents, and the selective use of scientific data do not make for effective conservation advocacy. Unfortunately, these tactics have characterized both sides of the sustainable-use debate. This book is no exception, and does a disservice to proponents of sustainable use and to our Cuban colleagues. ■

John G. Robinson and John Thorbjarnarson are at the Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, New York 10460, USA.

New in paperback

A History of the Mind: Evolution and the Birth of Consciousness

by Nicholas Humphrey

Copernicus, £7.99

Making Us Crazy: DSM — The Psychiatric Bible and the Creation of Mental Disorders

by Herb Kutchins & Stuart A. Kirk

Constable, £14.99

"This is a serious and well-documented study, which casts serious doubt on the touted scientific status of DSM categories. It is also readable, although Kutchins and Kirk's preoccupation with the day-to-day minutiae of the politics of

Seeds that never grew in Sweden

Linnaeus: Nature and Nation

by Lisbet Koerner

Harvard University Press: 1999. 298 pp.

£24.95, \$39.95

Harriet Ritvo

Linnaeus has long ceased to be a household name, even in scientific circles. During the nineteenth century, the twin enterprises of taxonomy and nomenclature, with which his name is associated, lost the position they had held during the Enlightenment at the cutting edge of botanical and zoological research. Further, as a Swede who published in Latin, Linnaeus lacks a modern constituency of any size. But his steady stream of taxonomic works made him an international scientific celebrity in a period when there was enough distinguished competition to lend enduring interest to his eminence.

Most Linnaeus scholarship has, understandably, focused on the work that inspired this contemporary renown. *Linnaeus: Nature and Nation* offers something different. It is neither a conventional biography nor a reinterpretation of Linnaeus's best-known scientific accomplishments, although it includes elements of both.

Instead, in a series of linked essays, Lisbet Koerner repositions Linnaeus primarily as a Swede rather than as a member of an international intellectual community. She emphasizes his deep family roots in the Swedish church and countryside, rather than his links to the larger world. After all, he spent only a few years of his long life outside his native land. He sent his disciples on dangerous voyages to the remote shores of Australia, Africa and Asia; for his own taste of the exotic, he headed north to Lapland.

Koerner shows that Linnaeus's professional preoccupations focused as much on

naming may dispose some psychiatrists to see in this a case of ancient obsessional disorder. It is certainly sobering to discover just how the terms we take for truth have come into currency."

Roy Porter, *Nature* **389**, 805–806 (1997)

Pluto and Charon: Ice Worlds and the Ragged Edge of the Solar System

by Alan Stern & Jacqueline Mitton

Wiley, \$19.95

Mathematical Mysteries: The Beauty and Magic of Numbers

by Calvin C. Clawson

Perseus, \$17



A Swede in Lapland: Linnaeus's scientific expeditions took him to Finland ...

local politics, academic and otherwise, as on the debates of foreign naturalists. He spent most of his career as a professor at Uppsala University, not far from Stockholm, where he wrote his students' dissertations (evidently an expected service) and revitalized the botanic garden.

He cut an eccentric and sometimes extravagant figure on the local scene. He led groups of up to 300 people — auditors as well as students, women as well as men — on collecting expeditions in the Uppsala countryside. These excursions were ultimately stopped by the university rector, who objected to them as both distracting and militaristic (Linnaeus imposed formal discipline and required participants to wear a uniform of his own devising). He supervised the careers of his former students, often finding them positions at home if they were fortunate enough to return from their travels — though, according to Koerner, Linnaeus was a difficult mentor and these efforts seldom led to mature professional friendships. He arranged for his son and namesake to inherit his own chair at Uppsala.

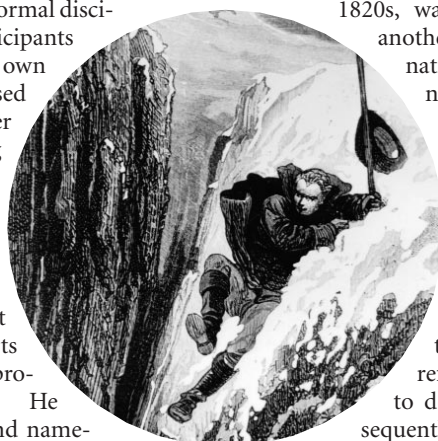
National issues also engaged Linnaeus, a

fact that Koerner emphasizes by using Swedish historical references (for example, to "the Great Northern Wars with Russia" and "the Era of Freedom"). She argues that some activities normally understood simply as part of Linnaeus's research gain significance when interpreted in the light of his ambitions for the emerging modern nation of Sweden. He wished to compensate for Sweden's limited ability to participate in European global expansion by naturalizing or domesticating some of the beneficial things found in other parts of the globe. He supported, along with many of his countrymen, cameralist economics (already old-fashioned in the world of Adam Smith) which aimed at national commercial isolation and self-sufficiency.

He particularly feared the haemorrhage of Swedish treasure to pay for tropical luxuries. Thus he looked for ways to supply consumer desires by more imaginative and efficient exploitation of native resources. As Koerner puts it, "He hoped to ride elks, write with swan feathers, and read by the light of seal-fat lamps". And if there were desires that could not be fulfilled in this way, Linnaeus hoped to persuade valuable tropical plants to adapt to his cold northern climate. He optimistically asserted that most Mediterranean plants could be acclimatized in Scandinavia, and that, if moved very gradually, perhaps over a century, even the fruits of India might be harvested in Uppsala. Guided by this theory he attempted to grow such crops as cacao, rice, coffee, sugar cane and pistachios in his botanic garden, although the results were invariably disappointing.

Perhaps because these failures had a political or economic dimension, Linnaeus' reputation suffered its swiftest decline in Sweden. Elsewhere, he dwindled gradually into a figurehead, the object of mechanical respect which he retains to this day. But in Sweden, where he was known as more than a systematist, things have been more volatile. If his post-mortem eclipse reflected his national political activities, his re-emergence as a Swedish hero, beginning in the

1820s, was due to politics of another kind. Romantic nationalists made Linnaeus the object of cult worship, which ultimately became associated with racist ideologies. By the 1920s, Linnaeus had become so closely identified with conservative politics that Swedish socialists refused to have anything to do with him. He consequently suffered a second



... where he came close to falling into a crevasse.

eclipse, at least in popular esteem. A glance at Koerner's bibliography suggests that a substantial proportion of scholarly work on Linnaeus continues to be published in Swedish, a fact that corroborates one of her central assertions. If all politics is local, all science is, too.

Harriet Ritvo is in the History Faculty, Massachusetts Institute of Technology, E51-285, Cambridge, Massachusetts 02139, USA.

An uncertainty principle for geometry

The Symbolic Universe: Geometry and Physics 1890–1930

edited by Jeremy Gray

Oxford University Press: 1999. 304 pp. \$105

Peter T. Landsberg

While physicists regard mathematics as a superb investigative tool, for the mathematician it is an end in itself. If, as in this book, the investigation is restricted to the relationship between geometry and physics, the situation is the same, except that a considerable variety of geometries exists: Euclidean, Riemannian, affine, and so on. So is there anything more of interest to be said about the relationship between geometry and physics?

The answer is a resounding 'yes', for Euclidean geometry was long regarded as the only possible geometry. However, around 1823, Nikolay Lobachevsky and János Bolyai dispensed with Euclid's axiom about parallel lines, so that in their geometry the angles of a triangle sum to less than 180 degrees. These new geometries suggested that Euclidean geometry was not 'obviously' the geometry of the Universe. In fact, this is a rather superficial view, as Einstein noted in his 1936 paper "Physik und Realität". Here he implied that centuries of Earth-bound measurements in agriculture and elsewhere made people think, wrongly, that Euclidean geometry was the only possible geometry.

If we are to rid ourselves of the Euclidean preconception, we must adopt a more philosophical approach and scrutinize the assumptions that had become quietly integrated into physics. Someone who dealt with just this problem was Hermann von Helmholtz. He remarked in one of his popular lectures, delivered in Heidelberg in 1870: "If it were useful for any purpose, we might with perfect consistency look upon the space in which we live as the apparent space behind a convex mirror with its shortened and contracted background ... Only then we should have to ascribe to the bodies which appear to us to be solid ... corresponding distensions and contractions, and we should have to change our system of mechanical principles entirely; for even the proposition that every

point in motion, if acted upon by no force, continues to move with unchanged velocity in a straight line, is not adapted to the image of the world in a convex mirror. The path would indeed be straight, but the velocity would depend upon the place."

Helmholtz's lecture was called "On the origin and significance of geometrical axioms", a topic covered in this collection edited by Jeremy Gray, and which Gray mentions at the end of part I of his book.

We can link Helmholtz's illuminating observation, and its early hint at curved space-time, introduced later, with a remark made by Roger Highfield and Paul Carter in their 1993 book *The Private Lives of Albert Einstein* (Faber & Faber). They point out that Mileva Maric and Einstein read Helmholtz's works together in 1899, four years before their marriage. Perhaps Helmholtz's writings inspired Einstein's relativity theory.

Such general observations, however, are rare in this book, which emphasizes the geometrical aspects. Thus, part I includes Heinrich Hertz's efforts to 'geometrize' mechanics and Henri Poincaré's views on geometry, while part II deals with the work of Hermann Minkowski and David Hilbert on relativity. Hilbert was much influenced by variational principles as well as by his axiomatic method; however, these principles were judged by the physicists of the period as unsuitable for relativity. Even so, Hilbert's broad scope and his understanding of the theories of physics are astounding.

Part III of the book looks at absolute differential calculus, congruences, Pauli matrices and other more specialized topics. However, there is also the more easily understood story by Umberto Bottazzini of how Gregorio Ricci-Curbastro failed to receive the Royal Prize for Mathematics for 1901, which had been announced by the Accademia dei Lincei. This chapter tells us how, after three years and in spite of strong applicants, the prize was postponed until 1905. Ricci-Curbastro never won it.

The reader can easily dip into this book, since it presents nine authors who seem to have been given rather wide terms of reference. Because of this, it must have been hard for an editor to pull their stories together and to extract clear lessons from them which could inspire the typical reader. I'm afraid Gray has not succeeded in achieving this, which is a shortcoming in what was otherwise a worthwhile undertaking. However, if all else fails in this respect, we can at least take away, and remember, one older, but valuable, idea. This is "Einstein's (1924) uncertainty principle" (my term), which says that "insofar as the propositions of geometry apply to reality, they are not certain; insofar as they are certain, they do not apply to reality". ■

Peter T. Landsberg is in the Department of Mathematics, University of Southampton, Southampton SO9 5NH, UK.

A fraternal view of Russian fauna

Animals in Nature

by Vladimir Smirin & Yuri Smirin

(translated by G. P. Harper)

Russian Nature Press: 1999. 307 pp.

Mikhail Mina

Many people think that if a book is 'scientific', it must be boring and end with a list of references. By these criteria, *Animals in Nature* is not a scientific book, even though the Russian version of the book, published in 1991, is valued by professional zoologists. *Animals in Nature* is intended primarily for people who like animals and want to know more about them.

Now the English version of the book is available. Geoffrey Harper's translation is faithful to the peculiarities of the authors' writing style, and the English edition includes maps of places mentioned in the text and notes explaining common Russian names of animals and geographical places.

The authors, brothers Vladimir and Yuri Smirin, are zoologists and wildlife artists, and have travelled through many parts of the former Soviet Union — the Far East, Chukotka, Kamchatka, the Komandor Islands, Kazakhstan, the Caucasus, and the White and Caspian seas. They observed rodents and antelopes, seals and chamois, wolves and Arctic foxes, whales and walrus. The notes of a keen observer of animals' lives are enhanced by excellent drawings of the animals. Some of the drawings can be viewed on the Internet (<http://www.rusnatpress.org.uk>).

Books on animals are often lavishly illustrated with glossy photographs. Such photographs may seem more exact and more objective than drawings, but, in the words of the authors, "Photographs never achieve the same organic unity with the text as can the author's own drawings, so



In the eye of the beholder: A chipmunk drawn by Vladimir Smirin.

that viewing the photographs and reading the text are generally separate acts ... While photography or film is essential for analyzing movement, drawings are better for a living portrayal in which the author can record personal impressions and experience, and represent what was actually seen."

Stories about animals written by naturalists and illustrated with the authors' drawings form a distinctive genre of fiction merging with wildlife art. This genre was and is very popular in Russia. Several generations of Russian naturalists were raised on the books of Ernest Seton-Thompson and Alexandr Formozov. The book by Vladimir and Yuri Smirin deserves the same ranking.

The abundance of wildlife photography books available is most welcome, but it would be a pity if they supplanted the works of wildlife artists who provide deep insight into animals' lives. ■

Mikhail Mina is in the Laboratory of Postnatal Ontogeny, Institute of Developmental Biology, Russian Academy of Sciences, Vavilov St. 26, Moscow 117334, Russia.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 21 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1998 and published at least four separate numbers by the end of May 2000.
- Journals covering any aspect of science are eligible, although those dealing with

clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.

- Frequency of publication must be at least three times a year.
- The main language must be English.
- The deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Isobel Flanagan, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)20 7843 4542.
e-mail: i.flanagan@nature.com

Venture funding for new ideas

Explanations for anomalies should be sought with an open mind.

Jayant V. Narlikar

For centuries, kings in India had a tradition of rewarding excellence in all creative fields, such as literature, music, fine arts and craftsmanship. The needy scholar had only to approach the royal court, demonstrate his creativity and be assured of financial support. However, from the seventeenth century onwards, as science began to make its presence felt in Europe, the Indian potentates did not appreciate its potential. The lack of royal patronage is one of the many suggested causes of the underdevelopment of science in India. This royal apathy came at a heavy price, as one of the reasons for the colonial dominance of India was the lack of growth of science and technology in the country.

In Europe, there were patrons for science among the royalty and rich aristocracy. Some scientists, such as Edmund Halley, Henry Cavendish or Charles Darwin, were well-to-do and could support their own research. But this trend of the eighteenth and nineteenth centuries began to change in the twentieth century, as scientific research became more and more expensive. Ernest Rutherford's pioneering experiments on the atomic nucleus were performed with apparatus costing around £100. Modern high-energy accelerators cost billions of dollars, for in order to delve deeper into the ultimate structure of particles, probes of higher and higher energy are required. At the other end of the size spectrum, astronomers need more and more expensive infrastructure to observe the Universe to fainter and farther limits. It is a far cry from Galileo's one-inch telescope of 1609, designed by himself, to the Keck Telescope of 10 metres' aperture nearly four centuries later.

Once such expensive facilities are created, often using the resources of several nations, it becomes essential to make the most efficient use of them. So applications to use the equipment are invited and judged by a peer-review process. The reviewers are conscious of the fact that the proposed experiment should

The extreme caution exercised by reviewers allows only the very safe proposals to get through.



A king's ransom? Royal patronage is no longer a reliable source of scientific funding.

advance the frontiers of knowledge, but at the same time it should be one for which the facility is essential. Thus proposals that are not scientifically well motivated or which can be carried out using other less sophisticated facilities are turned down.

All this looks reasonable. Yet the extreme caution that must be exercised by the reviewers allows only the very 'safe' proposals to get through. Here, safety is judged through the prevailing paradigm. What is expected under the existing paradigm? Does the proposal advance the existing knowledge one step further? If so, it is a good proposal. But these tactics do not allow for the unexpected, for the anomalous, for something that requires a radical modification of the existing paradigm.

I recall a story told by Subrahmanyan Chandrasekhar, the Nobel prizewinning astrophysicist. During a press conference in the 1930s, about the proposal to build the 200-inch telescope on the Palomar Mountain, Edwin Hubble and Arthur Stanley Eddington were asked what they expected to find with the new telescope. Their reply was: "If we knew the answer, there would be no purpose in building it."

Indeed, astronomy has advanced when-

ever the unexpected occurred, when a paradigm held sacrosanct by the majority of astronomers was shaken up. Such shifts occurred when it was realized that the Solar System is not at the centre of the Milky Way Galaxy, and again when it turned out that a good many nebulae are not part of our Galaxy but are galaxies in their own right lying far beyond the Milky Way. These paradigm shifts were not easy to achieve. Fritz Zwicky's ideas in the 1930s that clusters of galaxies may contain a lot of dark matter, or that the clusters may serve as gravitational lenses, were largely ignored at the time and we had to wait four decades to see them gain acceptance. Quasars and pulsars were both unexpected and anomalous in terms of the astrophysics of the time.

Closer examination of apparent anomalies can either show them to arise from artefacts, whereupon they cease to be important, and thus strengthen the existing paradigm, or the anomaly may force a change or modification of the paradigm. Either way, resolving the problem helps to develop the subject. Today, with large funds at stake, new facilities are proposed, designed and used with such a focused outlook that there is the danger of anything 'out of the way' being completely missed. Unlike Hubble and Eddington, today's proposers for a new facility begin by stating clearly what they expect to find with it.

Where does that leave the unexpected? Could the funding agencies and powers-that-be not reserve, say, ten per cent of their resources for venture ideas, for investigations of anomalies, for exploring alternatives to the existing paradigms? Provided, of course, that their originators have well-established credentials. Otherwise, science will be forced further away from what was once a free and open enquiry into the unknown, towards an exercise limited to confirming what is known.

Jayant V. Narlikar is at the Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Pune 411 007, India.



Regenesis

Transcript of the last cybercast from Fox-Warner reporter Daniel Lundgren.

Cynthia Ward

LUNDGREN in a hallway bustling with suited men and women: This is Dan Lundgren, Fox-Warner NewsNet, in Paris. I'm at the World Health Organization Conference on Genetic Therapy, where doctors and scientists are expressing alarm at the latest fashion trend. 'A fashion trend?' you might say. Medicine has nothing to do with style: but a medical breakthrough has become just another fashion statement.

Cut to a young woman on-stage, playing electric guitar in front of a thunderous retro-industrial band. Cut to LUNDGREN back-stage with the woman: We're speaking with Marie Durand, lead guitarist of Jack-hammer. Marie, many people would say you're abusing genetic therapy, desecrating a medical miracle for the sake of *haute couture*.

DURAND: They are fools. It has nothing to do with style.

Her hands rise into sight, gesturing. They seem too large for her slight size.

Fashion is transient. Art is eternal.

Close-up of one hand: The fingers are too long for the palm, and there are too many of them. Six.

I cut off my fingers. I would never do such a thing, except for art.

Cut to a bandaged hand, a stump without

fingers or thumb. Time-lapse video shows five bumps peeping through the bandage, and the bandage bulging along the outside edge of the palm. Then the palm is bare, fingers and thumb growing — growing too long — and a second thumb swelling from once-raw flesh.

DURAND [voice-over]: I cut off my fingers only to become the greatest guitarist in the world.

Cut to a small tailless lizard, sprouting a tail in time-lapse.

LUNDGREN [voice-over]: Once genetic engineers decoded the lizard's ability to regenerate a lost tail, human limb regeneration was no longer a fantasy. But no one realized the uses to which it would be put.

Cut to a beautiful Congolese woman with one brown eye, one green eye and, in the centre of her forehead, one blue eye.

[voice-over] Are we taking regeneration too far?

Cut to LUNDGREN, alone: One thing is clear: genetic therapists are giving plastic surgery a whole new meaning.

Cut to LUNDGREN with an Indian man in a Western suit tailored to reveal a long, supple, black-haired tail: Dr Charaka Ashok is a genetic engineer participating in the WHO Conference. Dr Ashok, many would claim you've made a mockery of your profession by giving yourself a monkey's tail.

ASHOK: They speak from ignorance. I haven't introduced a single gene from another species into my body. My tail is from the human genetic code — from 'junk' DNA. Junk genes are fossils: dormant genes inherited from our prehuman ancestors. Selective activation of the lizard gene and other fossil genes can stimulate the growth of tails, fangs, body fur and other physical features not normally part of the human body.

LUNDGREN: I see. But why would a distinguished scientist grow a tail?

ASHOK: It is a demonstration of truth. How better to manifest the fact of evolution? How better to refute fundamentalists who claim man was created by God? And refutation is more necessary than ever, with so many fundamentalists seeking to impose their ignorance on others. Religious terrorists are murdering genetic engineers and bombing gene-therapy clinics. Fanatics are clashing the world over.

Jump-cuts: screaming protesters; bombed-out buildings; battlefields.

In India, extremist Hindus, Muslims and Sikhs are in violent conflict with one another. And an apocalyptic Hindu sect has declared a man the latest avatar of the god Vishnu.

Cut to: long shot of hillside in India crowded with people, surrounding a man in richly coloured traditional clothing. The man is mounted on a white horse. He holds aloft a sword bathed in holographic flames. He has four arms.

ASHOK [voice-over]: According to Hindu belief, the next incarnation of Vishnu will be the tenth — and last. The last avatar, Kalki, heralds the end of the Kali-Yuga, the Iron Age — the end of the world as we know it. The extremists believe Kalki is here and it is their duty to bring the world to an end.

Cut to LUNDGREN and ASHOK.

LUNDGREN: Dr Ashok, I find it hard to believe a mob of poor, ignorant fanatics can destroy the world.

ASHOK: I was in the University of Washington graduate biotech program with Kalki when he was two-armed Sunesh Bannerjee. Kalki is a brilliant virologist and genetic engineer. He has the knowledge and the ability to kill millions.

Cut to LUNDGREN alone: There you have it. Regenerative changes that look like fashion are, in fact, acts of faith. Guitarist Marie Durand believes adding a second thumb to each hand will make her a great artist. Dr Charaka Ashok believes his tail is evidence of the truth of Darwinism. And it seems a brilliant, four-armed scientist may believe himself a god destined to destroy the world. Tomorrow night at 9 p.m. Eastern, we investigate Dr Sunesh 'Kalki' Bannerjee. With Fox-Warner NewsNet, this is Dan Lundgren.

Last private communication from LUNDGREN: Why'd you ignore my e-mails and voice-mails telling you to cancel upload of my WHO report? I told you I'm too sick to move, never mind investigate Kalki. I don't remember drinking any local water, but I must've forgotten to tell somebody to hold the ice for a Coke. I've got dysentery from hell. And India's hotter than hell and I can't get anyone to fix my air conditioning. This is supposed to be a four-star hotel, but I haven't been able to reach a staff member all day. They're not answering the phone, you're not answering the phone — oh. Oh, God, no —

Cynthia Ward is a freelance writer living in Seattle. She has stories in New Amazons (DAW Books) and Bending the Landscape: Horror (Overlook Press).



Perfect use of imperfection

Hansjörg Schild and Hans-Georg Rammensee

When making proteins, cells opt for high speed at the cost of large amounts of waste. But the waste-recycling process has a happy side effect — it gives the immune system a head start in its battle against viral invaders.

Every production process, be it industrial or cellular, aims at high efficiency. To achieve this end, the management must choose between high output at the cost of high waste, or slow, precision manufacturing with minimal error. Protein production in cells seems to occur by the high-speed, high-waste route. In mammalian cells, however, the waste may have an important function in the fight against viral intruders, as shown by Schubert and colleagues¹ and by Reits and co-workers² on pages 770 and 774 of this issue.

New proteins are assembled in cells by the lining up of amino acids, one after another, in a particular order. The newly synthesized amino-acid string undergoes complex folding and is frequently decorated with small attachments. Only then can it start to function — for example in cellular housekeeping or signalling.

The sequence of a protein is determined by the DNA archive that encodes that protein. Generating a protein involves copying (transcription) of the DNA into an RNA 'template' — the working copy — and

translation of the RNA, by ribosomes, into protein. This system is prone to error. Duplication of the DNA archive is carefully monitored by cells, but once the DNA is transcribed to RNA, production is stepped up and the manufacturing process is no longer carried out with the same degree of accuracy. The synthesis of some wayward proteins may be a waste of cellular resources, but it is not as damaging as a genetic mutation, which may alter not only a cell but also all of its progeny.

Schubert *et al.*¹ reveal just how wasteful a cell can be when making proteins. The authors start their investigation by inhibiting the main protein-degrading engine in cells — the 26S proteasome. Inhibiting the proteasome in this way leaves higher levels of newly produced cellular proteins intact. Schubert *et al.* show that at least some of the proteins are tagged with a small molecule called ubiquitin, which singles proteins out for proteasomal destruction. The assumption is that, as these surplus proteins are destined to be degraded in the proteasome, they are probably faulty in some way — the results

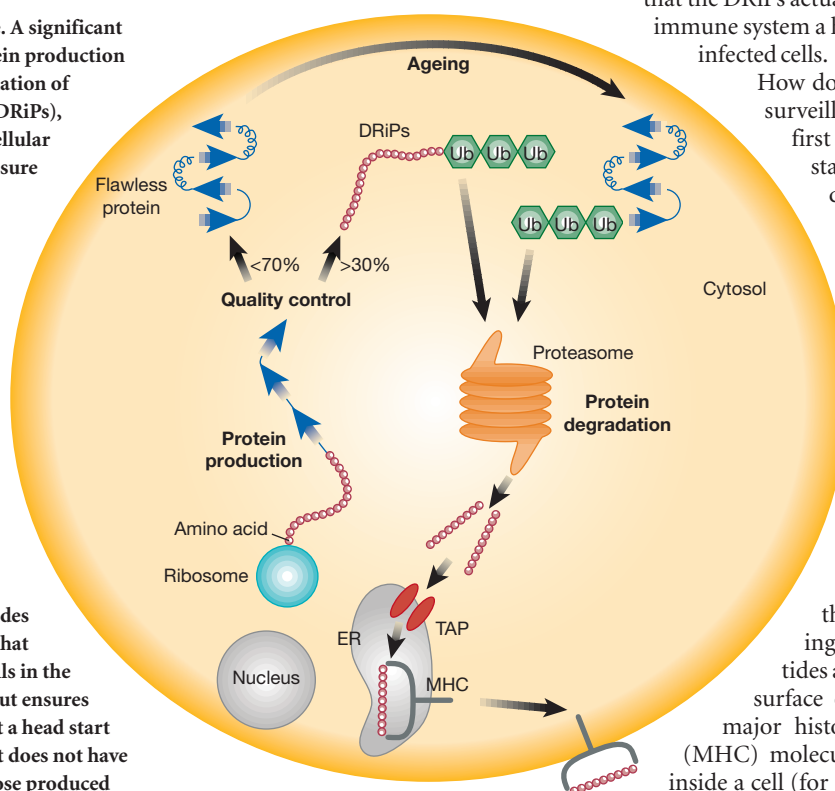
of rapid but error-prone protein synthesis (Fig. 1). The authors proposed the existence of these faulty proteins — called defective ribosomal products, or DRiPs — four years ago³ on the basis of observations of the rapid generation of peptide fragments from apparently metabolically stable proteins. Surprisingly, Schubert *et al.* now show that the failure rate of protein synthesis (that is, the rate of DRiP formation) in the different cell types that they studied exceeds 30%.

However, the cell's production management is not as negligent as this figure might suggest. First, every faulty protein is recycled into amino acids and reused for the synthesis of new proteins. Second, as Schubert *et al.*¹ and Reits *et al.*² describe, mammalian cells use these waste products to provide advance warning about the proteins currently in production (Fig. 1). This gives the immune system a helping hand. For example, a cell that is infected by a virus will be producing viral proteins as well as its own. The immune system needs to recognize that the cell is infected, and destroy it. And it seems that an important side effect of the DRiP system is that the DRiPs actually give the T cells of the immune system a head start in recognizing infected cells.

How does the immune system's surveillance system work? It is first important to understand that the proteasome does not only degrade faulty proteins; it also destroys flawless proteins that have simply reached the end of their working lives.

Before the revelations of Schubert *et al.*¹ and Reits *et al.*², it was assumed that immune cells detect infected cells in the following way. Flawless, aged, foreign proteins are chopped up in the proteasome, generating peptides. These peptides are then displayed on the surface of the cell by so-called major histocompatibility complex (MHC) molecules. Peptides produced inside a cell (for example, viral peptides) are displayed mainly by MHC class I

Figure 1 A use for protein waste. A significant amount (30% or more) of protein production is faulty and results in the formation of defective ribosomal products (DRiPs), as shown by Schubert *et al.*¹. Cellular quality-control mechanisms ensure that DRiPs are immediately removed by becoming tagged with a small molecule called ubiquitin (Ub) and targeted to the proteasome, which chops the DRiPs up into small peptides. Samples of the peptides are transported by the transporter associated with antigen presentation (TAP) into the endoplasmic reticulum (ER), as suggested by Reits *et al.*², loaded on MHC class I molecules and transported to the cell surface for recognition by T cells. Peptides from fully functional proteins that have aged are presented to T cells in the same way. But the DRiP short cut ensures that the immune system can get a head start on fighting infection, because it does not have to wait for proteins (such as those produced in virus-infected cells) to age.



molecules. (By contrast, peptides derived from extracellular pathogens, such as bacteria, are displayed mainly by MHC class II molecules.) The proteasome is found in the main body of the cell, but the MHC molecules are first found in the membrane of an organelle called the endoplasmic reticulum (ER), so the peptides too need to get into the ER. A protein called TAP (for transporter associated with antigen processing) transports peptides into the ER. The peptides are then loaded onto the MHC molecules, which travel with their cargo to the cell surface. Circulating T cells recognize the MHC-peptide complexes, and spark off an immune response to the infected cell, resulting in its death.

As far as we knew, this mechanism worked only for flawless but aged proteins. However, peptides that must be derived from abnormal translation events have also been found to be expressed by cells, in such small amounts that their expression could be detected only by T cells⁴⁻⁷. Such erroneous translation events presumably contribute to the pool of DRiPs. So, it seems that in fact there is a shortcut from translation to degradation. The cell does not have to wait for flawless proteins to be produced, to age, and then to be degraded into MHC-presentable peptides. Instead, it can make use of the waste products generated by the cell's fast but faulty production system.

Schubert *et al.*¹ and Reits *et al.*² adopt two very different approaches to confirm that the cell does indeed display peptides from faulty proteins on MHC molecules. Schubert *et al.* observe that the maturation of MHC class I molecules (which occurs as they move from the ER to the cell surface) is correlated directly with the rate of production of DRiPs and their degradation by proteasomes. Reits *et al.*, by contrast, take the mobility (lateral diffusion in the ER membrane) of fluorescently labelled TAP molecules as an indicator of peptide transport and thus of intracellular peptide levels. These authors find that TAP's mobility decreases immediately after they introduce peptides into the cell or within the first few minutes after virus infection — presumably because TAP is then actively transporting peptides into the ER, rather than 'looking' for peptides. Newly synthesized proteins, probably DRiPs, cause the reduction in TAP mobility: Reits *et al.* show that protein synthesis and proteasome activity (which result in, respectively, DRiP synthesis and degradation into peptides) are essential for the decrease in TAP mobility.

So, the immune system is not condemned to wait until properly folded foreign proteins have been degraded and their fragments presented for recognition. Instead, by using the DRiP short cut, the immune system can initiate countermeasures while the foreign proteins are still being produced. Abnormal

protein translation increases both the speed of generation of potential MHC class I cargoes — a feature important for controlling virus spread — and, perhaps, the range of cargo peptides. Waste not, want not. ■

Hansjörg Schild and Hans-Georg Rammensee are in the Department of Immunology, Institute for Cell Biology, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.
e-mail: hansjoerg.schild@uni-tuebingen.de

Nonlinear dynamics

Chaos in space and time

J. P. Gollub and M. C. Cross

Many aspects of nature are essentially unpredictable over the long term, even when quantum effects are completely absent. A frequent cause is chaotic dynamics, which can arise when the rates of change of important dynamical variables depend in a nonlinear fashion on the variables themselves. Famous examples are the unpredictable behaviour of the weather¹ and of populations of organisms². Chaotic temporal dynamics has now been documented in practically every scientific discipline — the fascination for both scientists and the public is that chaos helps us understand what can be predicted about the systems we study, and what cannot. But our understanding of chaos in most systems found in nature, which are complex in both space and time, is relatively limited. On page 733 of this issue, Egolf *et al.*³ analyse the chaotic space and time variations of a heated fluid in a way that could eventually be applied to natural systems such as the weather.

The techniques of nonlinear dynamics are well developed, but its impact has been largely confined to phenomena in which there are only a few important time-dependent quantities, such as the position and velocity of a swinging pendulum. Unfortunately, this excludes a vast range of important problems in which the behaviour at one point in space can be quite different (though statistically similar) to that at another location. For example, heating a fluid from below⁴ — as the Earth's atmosphere is warmed by contact with the Earth — causes upwelling at some locations and descending flow elsewhere. This convective behaviour can be highly organized and predictable if the heating isn't too great, but the flow pattern becomes complex and apparently unpredictable as the heating is increased, for instance in the surface layer of the Sun. This phenomenon, frequently called 'spatiotemporal' chaos because the complex dynamics involve variations in both space and time, is the subject of Egolf *et al.*'s study.

The traditional approach to studying

1. Schubert, U. *et al.* *Nature* **404**, 770–774 (2000).
2. Reits, E. A. J., Vos, J. C., Grommé, M. & Neeffjes, J. *Nature* **404**, 774–778 (2000).
3. Yewdell, J. W., Anton, L. C. & Bennink, J. R. *J. Immunol.* **157**, 1823–1826 (1996).
4. Boon, T. & Van Pel, A. *Immunogenetics* **29**, 75–79 (1989).
5. Elliott, T., Bodmer, H. & Townsend, A. *Eur. J. Immunol.* **26**, 1175–1179 (1996).
6. Malarkannan, S., Horing, T., Shih, P. P., Schwab, S. & Shastri, N. *Immunity* **10**, 681–690 (1999).
7. Mayrand, S. M. & Green, W. R. *Immunol. Today* **19**, 551–556 (1998).

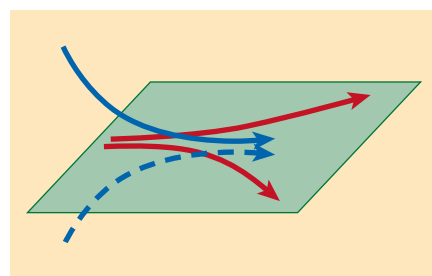


Figure 1 Spatiotemporal chaos. In a chaotic system, nearby trajectories separate exponentially along some directions in phase space (red arrows), while they converge along other directions (blue arrows). In low-dimensional chaos this combination leads to the familiar sheet-like fractal structures known as 'strange attractors'. By projecting the dynamics onto the most rapidly expanding directions, Egolf *et al.*³ can identify the dynamical processes important to chaos in very large systems.

nonlinear dynamical behaviour is to plot the system's dynamical variables as a multi-dimensional 'phase space' graph showing how the behaviour changes over time. A point in phase space represents one possible instantaneous state of the system, and a sequence of these points shows the time history of the system. For example, a simplified model of the Solar System consisting of the Sun and nine planets would require a phase space with as many as 60 dimensions (three position and three momentum coordinates for each body). In the case of a convecting fluid, a complete description of the flow pattern requires knowledge of the velocity and temperature at a very large number of locations, so the dimension of the phase plot is enormous (from thousands to millions, depending on the desired spatial resolution). As a result, the methods of nonlinear dynamics are cumbersome and progress has been disappointingly slow, even though many interesting examples of spatiotemporal chaos have been explored experimentally and numerically.

Egolf and colleagues³ have carried out



100 YEARS AGO

Mr. Herbert Spencer, exposing the various inconsistencies that occur so frequently in the ghost-stories of the savage races, says:—"How illogicalities so extreme are possible, we shall the more easily see on recalling certain of our own illogicalities. Instance ... that familiar absurdity fallen into by believers in ghosts, who, admitting that ghosts are seen clothed, admit, by implication, that coats have ghosts—an implication they had not perceived." It seems interesting to note that the same opinion was expressed about nineteen centuries ago by the Chinese philosopher, Wang Chung (circa, 27–97 A.D.), whose sceptic remarks on the traditions of all manners, handed down to his time in the Middle Kingdom, form a celebrated work named 'Lun Han' or 'Balance of Discussions'. It is curious to observe that Wang Chung himself is quite illogical in esteeming it just to suppose a ghost able to appear only divested: for, according to his own proposition, the soul exists only in blood and breath; while the body, though very closely connected with them during life, is, after death, as severed from them as the ever lifeless and soulless clothes; so that, should it be necessary for a ghost to appear divested, it would be equally so to appear disembodied at the same time. From *Nature* 12 April 1900.

50 YEARS AGO

There is a tendency nowadays to look unfavourably on any book about the production and supply of food. This is perhaps not so much because we are all acutely food conscious, but rather because we have come to regard food in the prosaic light of calorific content. That admirable principle, "a little of what you fancy does you good", can no longer be applicable in an age of restricted choice. Our appetites have had to be modified by the limitation of our resources. Necessity has become our master and the written word the means of persuading us that though we may no longer fancy what we eat, what we do eat may still be good for us because it contains so many units of energy. If the literature about food reflects in some degree the economic background of the period in which it is written, we, who live in 1950, may be excused if we prefer nostalgic memories and are disinclined to read about foods which we are forced to eat to sustain our bodies but which leave our minds unaffected by their dullness. From *Nature* 15 April 1950.

numerical studies of an accepted model of thermal convection. They use their results to reveal, for the first time, the mechanism by which the chaotic dynamics is generated. A key concept needed to understand chaos is its sensitive dependence on current conditions (Fig. 1). When the history of the system is plotted in multidimensional phase space, two trajectories or paths that differ slightly at some instant will separate exponentially fast from each other along certain directions (while approaching each other exponentially fast along other directions).

Understanding the dynamics means understanding this exponential growth, which can be characterized by parameters known as Lyapunov exponents. The number of Lyapunov exponents is equal to the dimensionality of the phase space. Egolf *et al.* determine the first 100 or so of these exponents for their convecting system, and show that the pattern of exponents is independent of the size of the system in the sense that doubling the system size causes new exponents to appear between the old ones. Furthermore, the largest exponents correspond to the creation of localized defects in the convection pattern — that is, places where the pattern of rising and falling motion of the fluid changes its topology — not unlike the appearance of defects in a disordered crystal lattice when it is deformed.

The implication of this work³ is that the origin of the unpredictable motion, at least in this particular form of spatiotemporal chaos, lies in what happens in small regions of space and over short timescales. These local changes in the organization of the

flow affect the surrounding regions in such a way that the entire future evolution is affected. This is something akin to Ed Lorenz's famous remark that the localized flapping of a butterfly's wings might change the weather dramatically over the entire world a few weeks later, except that here the localized events occur naturally; they are not external interventions. Although such sensitivity to localized fluctuations has never been confirmed as the source of the weather's unpredictability, it is apparently the origin of chaotic dynamics in thermal convection.

The methods used by Egolf *et al.* should apply to many other forms of chaos in spatially extended systems (physical, chemical and biological) for which reliable model equations are available, so that the key processes leading to the complex dynamics can be identified. Applications to areas as diverse as cardiology and atmospheric dynamics might be expected eventually. Moreover, it is not unreasonable to imagine that insight into the processes leading to unpredictability will also lead to progress in modifying or controlling the dynamics of these systems. ■

J. P. Gollub is in the Department of Physics, Haverford College, Haverford, Pennsylvania 19041, USA.

e-mail: jgollub@haverford.edu

M. C. Cross is in the Department of Physics, California Institute of Technology 114-36, Pasadena, California 91125, USA.

e-mail: mcc@caltech.edu

1. Lorenz, E. N. *J. Atmos. Sci.* **20**, 130–141 (1963).
2. May, R. M. *Nature* **26**, 459–467 (1976).
3. Egolf, D. A. *et al.* *Nature* **404**, 733–736 (2000).
4. Bodenschatz, E., Pesch, W. & Ahlers, G. *Annu. Rev. Fluid Mech.* **32**, 709–778 (2000).

Plant biology

Shattering developments

Günter Theißen

Imagine a world without fruits: there would be no apples and oranges, of course, but also no bread, no noodle and rice meals, and (oh my God!) no wine and beer. An improved understanding of the genetic control of fruit development should help us to design better fruits for various purposes. On page 766 of this issue¹, Liljegren *et al.* report advances which provide a starting point to do just that. At the same time, the authors' data give an exciting glimpse of the role of developmental control genes in fruit evolution.

Fruits mediate seed maturation and dispersal, and are derived from flowers. So they are unique to flowering plants, the angiosperms (as opposed to the gymnosperms, which produce unprotected, or

'naked', seeds). The model plant *Arabidopsis thaliana*, a tiny weed from the mustard family Brassicaceae, produces dry, dehiscent fruits — that is, they burst open at maturity to release the seeds. *Arabidopsis* fruits are composed of two valves (carpel walls) separated by a thin structure termed the replum^{1,2} (see Fig. 1 of the paper on page 767). At the valve–replum boundary, a narrow band of cells develops into the dehiscence zone. Late in fruit development, cells in the 'dehiscence zone' separate from one another, the valves detach from the replum, and the seeds can disperse.

Liljegren *et al.* have generated *Arabidopsis* plants in which two very similar genes, *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*), have both lost their

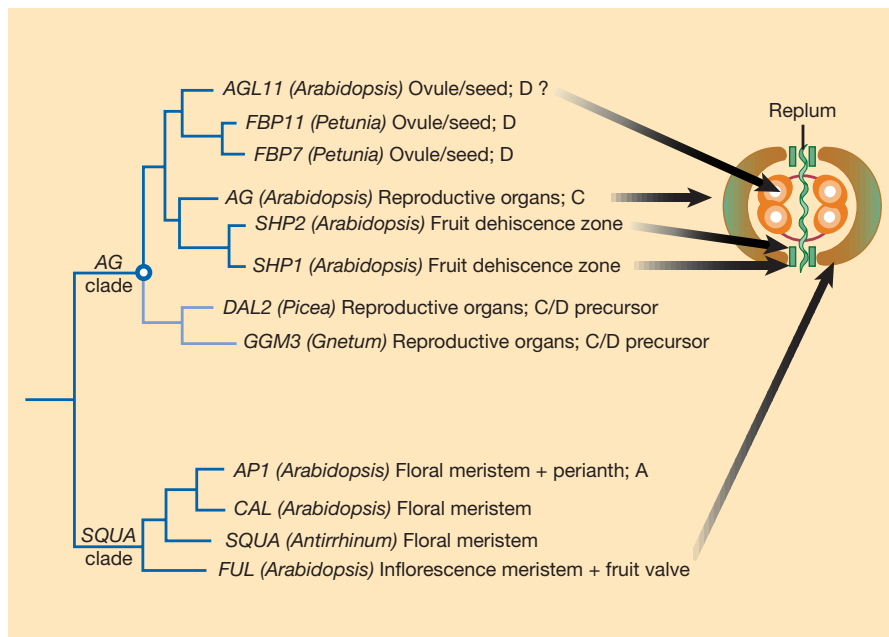


Figure 1 MADS-box genes and fruit evolution. The cladogram shows the relationship between some examples of MADS-box genes. A, C and D indicate gene functions involved in specifying the identity of particular parts of the flower: sepals and petals (A), stamens and carpels (C), and ovules (D)^{7,8,10}. *DAL2* and *GGM3* are genes from two gymnosperm species; all the others are from flowering (angiosperm) plants⁸. The separation of angiosperm and gymnosperm members of the AG clade occurred about 300 million years ago (dot at the base of the AG clade). *SHP1*, *SHP2* and *FUL* are the *SHATTERPROOF1* and *2* and *FRUITFULL* genes investigated by Liljegren *et al.*¹; *AG*, *AGAMOUS*; *SQUA*, *SQUAMOSA*; *AGL11*, *AG-like1*; *FBP7/11*, *FLORAL BINDING PROTEIN7/11*; *AP1*, *APETALA1*; *CAL*, *CAULIFLOWER*.

function. In mature fruits of these double mutants, the dehiscence zones are absent, so the fruits fail to break open. The authors have complemented their loss-of-function analyses by examining transgenic plants that express the *SHP* genes constitutively (that is, all the time and in all tissues of the plant), and also by using putative downstream targets of these genes as molecular markers to monitor the cellular differentiation of valve margins.

To cut a long story short, it turns out that *SHP1* and *SHP2* ensure correct development of the cells in the dehiscence zone and of adjacent cells at the valve margin. Liljegren *et al.* also report evidence that another gene, *FRUITFULL* (*FUL*), interacts antagonistically with the *SHP* genes during development of the valve margin; *FUL* is required for valve-cell differentiation and expansion after fertilization².

These findings are agronomically important because *Arabidopsis* is closely related to oilseed crop plants, such as canola (*Brassica napus* and *Brassica rapa*) in which premature fruit dehiscence — pod shatter — causes considerable yield losses. Knocking out *SHP*, or overexpressing *FUL*, could result in shatter-resistant canola plants. Such plants might be generated either by genetic engineering (possibly the fast way, but not fashionable in Europe at the moment), or by random mutagenesis and marker-assisted breeding.

The studies of Liljegren *et al.*¹ will also help us to understand fruit evolution. Some reproductive characters³, including fruit indehiscence, evolved independently several times within the Brassicaceae, for example in close relatives of the genus *Lepidium* (pepper cresses) (K. Mummenhoff, personal communication). Because loss of *SHP* function (which in other plants is not necessarily provided by two different genes) has little effect apart from its influence on fruit dehiscence, it may represent a simple mechanism by which indehiscent fruits originated from dehiscent fruits.

Moreover, *SHP1*, *SHP2* and *FUL* (formerly known as *AGL1*, *AGL5* and *AGL8*, respectively^{4,5}) belong to the family of so-called MADS-box genes, which encode gene transcription factors in plants, animals and fungi^{6–8}. MADS-box genes are involved in many developmental processes in the plant life cycle. But they became famous, at least among plant biologists, because some of them specify the identity of the different floral organs during flower development. These organs are the leaf-like sepals; the often colourful and attractive petals; the stamens (the male reproductive organs); and the carpels (the female reproductive organs), inside which the ovules develop into seeds.

The MADS-box gene family is composed of several defined gene clades — sets of genes that share a most recent common

ancestor not shared with any of the other MADS-box genes^{7–9}. *SHP1* and *SHP2* are members of a group of closely related genes termed the *AGAMOUS* (*AG*) clade^{7,8} (Fig. 1). This clade also includes the genes known to specify stamen and carpel identity during flower development (such as *AG* itself), and those that specify ovule identity¹⁰. Members of the *AG* clade have also been found in gymnosperms (Fig. 1), but not in ferns, implying that the first member of this clade arose 300–400 million years ago⁸. At that time, flowers and fruits were not yet established — flowering plants probably originated not much more than 200 million years ago.

The ancestral function of members of the *AG* clade was probably to specify the primary identity of reproductive organs⁸. So it seems that the *SHP* genes have a highly specialized function (control of dehiscence-zone development) in a structure (the fruit) that originated relatively late in evolution and which itself serves a quite specialized function (seed maturation and dispersal) (Fig. 1). This provides a pretty good example of the involvement of new genes, generated by gene duplication, sequence divergence and fixation within the *AG* clade^{7,8}, in the evolution of reproductive devices within the past 200 million years.

Another striking example of functional gene recruitment is provided by the *FUL* gene. This belongs to the clade of *SQUAMOSA*-like genes, whose members operate in specifying meristem — proliferative tissue — identity^{7,8}. *FUL* itself still works in determining meristem identity in inflorescences^{2,8}. But it was obviously recruited during the course of fruit evolution for its additional function in the fruit valves (Fig. 1).

We are just at the beginning of understanding the role of MADS-box genes in fruit development and evolution. Clearly, however, fruiterers, as well as florists and evolutionary biologists, owe a lot to this dynamic gene family.

Günter Theißen is in the Department of Molecular Plant Genetics, Max-Planck-Institute for Breeding Research, Carl-von-Linné-Weg 10, D-50829 Köln, Germany.

e-mail: theissen@mpiz-koeln.mpg.de

1. Liljegren, S. J. *et al.* *Nature* **404**, 766–770 (2000).
2. Gu, Q., Ferrándiz, C., Yanofsky, M. F. & Martienssen, R. *Development* **125**, 1509–1517 (1998).
3. Bowman, J. L., Brüggemann, H., Lee, J.-Y. & Mummenhoff, K. *Int. J. Plant Sci.* **160**, 917–929 (1999).
4. Savidge, B., Rounsley, S. D. & Yanofsky, M. F. *Plant Cell* **7**, 721–733 (1995).
5. Mandel, M. A. & Yanofsky, M. F. *Plant Cell* **7**, 1763–1771 (1995).
6. Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H. & Sommer, H. *Science* **250**, 931–936 (1990).
7. Theißen, G., Kim, J. T. & Saedler, H. *J. Mol. Evol.* **43**, 484–516 (1996).
8. Theißen, G. *et al.* *Plant Mol. Biol.* **42**, 115–149 (2000).
9. Doyle, J. J. *Syst. Biol.* **43**, 307–328 (1994).
10. Angenent, G. C. & Colombo, L. *Trends Plant Sci.* **1**, 228–232 (1996).

High-temperature superconductivity

Stripes defeat the Fermi liquid

J. Zaanen

Some 50 years ago, profound insights were obtained into the nature of the metallic and superconducting states of electrons in solids. According to Fermi-liquid theory, electrons in ordinary metals can be viewed as a non-interacting electron gas. Then, as their temperature is lowered, the electrons form Cooper pairs that are responsible for superconductivity in the Bardeen–Cooper–Schrieffer (BCS) theory. On pages 729 and 736 of this issue, Mook *et al.*¹ and Sharma *et al.*² announce the results of two independent experiments showing that the conventional theories of metals and superconductors are no longer universally valid. At the same time, both works are inspired by and add further credibility to an alternative picture of electrons in solids known as ‘dynamical stripes’.

The first hints of a challenge to the established paradigm came with the discovery of the high-temperature superconductors in 1987. Soon after their arrival, suspicions arose that something completely different was going on in these copper oxide superconductors³. But the conventional theories proved flexible and were defended quite successfully. Mook *et al.*¹ and Sharma *et al.*² have now used different experimental probes (neutron scattering and ion channelling) aimed at dissimilar properties of the copper oxides (spin and ion-lattice fluctuations) to disqualify, once and for all, these theories as an explanation for high-temperature superconductivity.

The state of electrons in normal metals should be understood as an extreme expression of the perpetual motions of quantum

physics. In Fermi-liquid theory, the metallic state of matter corresponds to a quantum gas of ‘quasiparticles’, which are like electrons in every way except that they no longer interact. The quasiparticles are fermions and have to obey the Pauli exclusion principle, which forbids any two fermions from being in the same state. This forces the electrons into states of high quantum kinetic energy and, remarkably, the fierce quantum motions wash away completely the inter-actions between electrons. Conventional superconductivity is a sibling of this state. In the presence of any residual attractive interaction the quasiparticles form pairs and, because these Cooper pairs are bosons (and so are not constrained by the Pauli principle), they immediately condense into a BCS superconductor.

These quantum gases are very simple and rather featureless. The spin and lattice fluctuations in the copper oxides seen by Mook *et al.* and Sharma *et al.* reflect a complexity of behaviour in the electron system that cannot be attributed to any gas, even a quantum gas of quasiparticles. A high degree of cooperativity is needed to explain these diverse

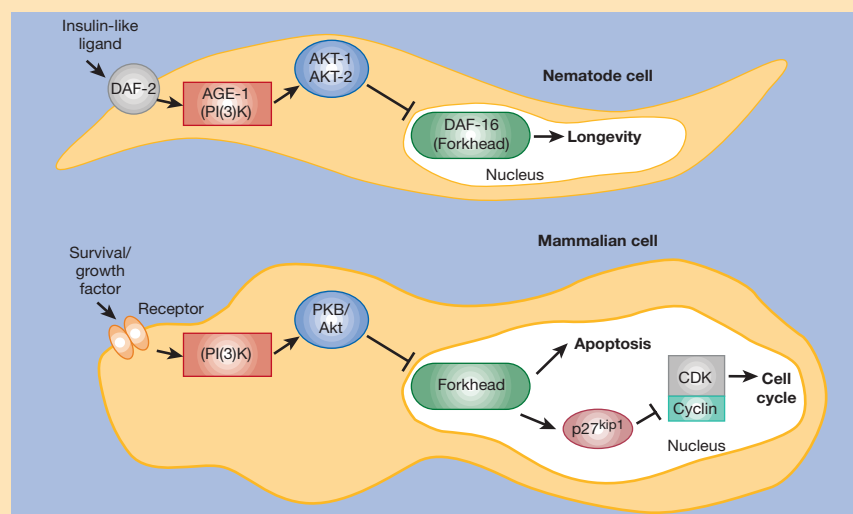
Signal transduction

An arresting tale

The external differences between species belie the fact that many of their genes, proteins and intracellular signalling pathways are very similar. So, for example, we can learn much about ourselves from studying such experimentally useful animals as the nematode *Caenorhabditis elegans*. Elsewhere in this issue (*Nature* **404**, 782–787; 2000), however, René H. Medema and colleagues turn the tables. They describe a new branch of a signalling pathway in mammalian cells that may also offer insight into an important developmental state in *C. elegans*.

The story starts with humanity’s fascination with extending life. For this reason, long-lived mutants of organisms such as *C. elegans* have received a lot of attention. By identifying the genes mutated in such organisms, we can get some idea about the proteins and signalling pathways involved in longevity. One pathway involves a cascade of enzymes known as kinases, culminating in the regulation of DAF-16 — a transcription factor of the so-called Forkhead family (see diagram). This pathway also regulates the developmentally arrested nematode larval state known as ‘dauer’, which occurs as a result of crowding and starvation.

At the cellular level, it is possible that, for a cell to stay alive, its death (by ‘apoptosis’) must be actively suppressed. In mammals, a pathway that represses apoptosis contains counterparts of many proteins from the nematode longevity pathway (see diagram; proteins that are conserved in the two pathways are shown in the same colours). This ‘survival’ pathway also ends with inhibition of members of the Forkhead family, resulting in



repression of genes of the death machine.

Medema *et al.* now show that this signalling pathway has another, possibly more important, function: it puts a brake on the cell-division cycle. The authors find that signalling progresses along the now familiar enzymatic cascade, up to the point at which Forkhead proteins are regulated. Then, the pathway forks. One line leads to suppression of cell death, while the other leads to cell-cycle arrest — the key molecular effect here being an increase in levels of the protein p27^{kip1}. This appears to occur through increased expression of the gene encoding p27^{kip1}, rather than through stabilization of the p27^{kip1} protein.

p27^{kip1} is a well known molecule: it inhibits the activity of important regulators of the cell cycle,

called cyclin/cyclin-dependent kinase (CDK) complexes. Increased p27^{kip1} levels would be expected to lead to cell-cycle arrest in the period before DNA replication occurs. Such a block is exactly what Medema *et al.* find when they activate Forkhead proteins in mammalian cells.

This new prong of a known signalling pathway may provide an explanation for the proposed role of defects in this pathway in cancer — a condition that is essentially characterized by the undermining of the finely tuned molecular network that ensures controlled cell division and finite cellular lifespan. This story may also have come full circle in explaining why, in nematodes, this pathway blocks larval development.

Bernd Pulverer

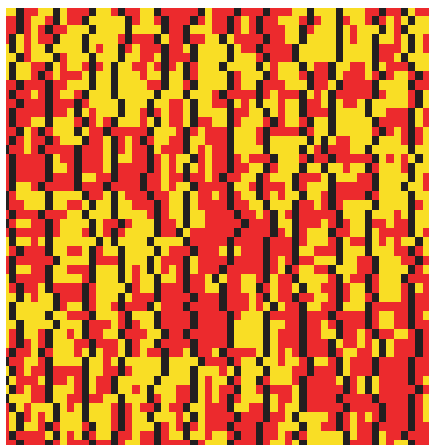


Figure 1 A snapshot of stripes in action. Fluctuating quantum stripes (lines of black dots) are shown moving through a quantum antiferromagnet (red/yellow background) from an imaginary time slice of a quantum Monte Carlo simulation. In the classical limit the background would be, say, uniformly red, corresponding to a simple antiferromagnet. Here the yellow patches represent spin fluctuations. The first conclusive evidence for the existence of stripes in high-temperature superconductors is given by Mook *et al.*¹ and Sharma *et al.*². A consequence of this work is that the conventional picture of ordinary metals and superconductors can no longer explain high-temperature superconductivity. (Unpublished figure, O. Y. Osman, J. Tworzydło and J. Zaanen.)

observations, and the best interpretations are found in the new theory of dynamical stripes. This increasingly popular theory refers to a state that is radically different from a Fermi liquid. In stripes, the electrons have been incorporated in complex, quantum fluctuating patterns and, although much is unclear, many now believe that superconductivity and stripes have a deep and profound relationship.

Owing to a spectacular series of discoveries in the past few years (see ref. 4 for a review), there now exists some understanding of the counterintuitive nature of this 'complex quantum matter'. In many-body systems (so-called quantum fields), quantum effects might turn out to be relatively modest at short-length scales, but increasingly important over larger distances. This is unlike few-body systems, where quantum mechanics invariably dominates at the shortest distances. Earlier neutron-scattering experiments⁵ suggested that this general field-theory mechanism was also at work in the copper oxide superconductors⁶. Because the long-wavelength quantum fluctuations involve small energy scales, one expects to easily suppress these, driving the state to its classical limit. Tranquada and co-workers⁷ accomplished this feat in 1995, and their frozen version of stripes allowed experimentalists to have a closer look at the nature of this complex electronic matter.

The static state of stripes is highly unusual (Fig. 1). It consists of antiferromagnetic (and presumably insulating) domains about a nanometre in width, separated by domain walls on which the charge carriers reside. In the static phase these domain walls, or stripes, condense in a regular pattern. Such phases were actually predicted theoretically^{8,9} (triggering the experimental search), and the current consensus is that they arise from competition between quantum kinetic energy and electron–electron interactions on the microscopic scale.

How can we understand dynamical (as opposed to static) stripes? This is the state associated with superconductivity, in which the system of stripes is disordered by long-wavelength quantum fluctuations. Figure 1 illustrates a popular view. It gives a snapshot at a fixed (imaginary) time from a quantum Monte Carlo calculation, and the overall quantum state should be viewed as a sequence of many of these pictures, in which the irregular features vary according to the quantum fluctuations. Yellow and red indicate opposite orientations of the stripe antiferromagnet: the background would be uniformly red in the classical limit. Although the stripes are still intact (lines of black dots), they undergo vigorous quantum motions. In fact, several theoretical groups are studying these stripe fluctuations in terms of simple string theories¹⁰.

The superconductivity community takes such pictures increasingly seriously, if only because of their effectiveness in guiding the experimental work. This does not mean that the problem of high-temperature superconductivity is solved. Over long distances the quantum fluctuations get truly out of hand, destroying the semiclassical picture presented above. Remarkably, in this regime there are similarities with a BCS superconductor, although these might not be as straightforward as some defenders of the conventional model want to believe. At the moment the relationship between stripes at short distances and BCS-like behaviour at large distances is an enigma. But I am an optimist and I take this mystery as prime evidence that there will be a glorious theory of physics waiting at the end of the journey. ■

J. Zaanen is at the *Instituut-Lorentz for Theoretical Physics, Leiden University, PO Box 9506, NL-2300 RA Leiden, The Netherlands.*

e-mail: jan@lorentz.leidenuniv.nl

1. Mook, H. A. *et al.* *Nature* **404**, 729–731 (2000).
2. Sharma, R. P. *et al.* *Nature* **404**, 736–740 (2000).
3. Anderson, P. W. in *The Theory of Superconductivity in the High T_c Cuprates* (ed. Treiman, S. B.) (Princeton Univ. Press, 1997).
4. Zaanen, J. *Science* **286**, 251–252 (1999).
5. Aeppli, G. A. *et al.* *Science* **278**, 1432–1435 (1998).
6. Sachdev, S. *Phys. World* **12**, 33–38 (1999).
7. Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. *Nature* **375**, 561–563 (1995).
8. Zaanen, J. & Gunnarsson, O. *Phys. Rev. B* **40**, 7391–7394 (1989).
9. Emery, V. J. & Kivelson, S. A. *Physica C* **235–240**, 189–192 (1994).
10. Zaanen, J. *Phys. Rev. Lett.* **84**, 753–756 (2000).

Functional genomics

Recognizing DNA in the library

Satish K. Nair and Stephen K. Burley

The transcription of DNA into RNA requires a battery of proteins, the precise complement of which depends on the circumstances. One group of such proteins is the basic helix–loop–helix (bHLH) family of DNA-binding transcription factors. So far, over 250 of these bHLH factors have been characterized; they regulate gene expression in various processes such as muscle development, circadian rhythms and cell growth¹. In a paper in *Chemistry and Biology*², Winston and Gottesfeld use a novel combinatorial approach to provide new information about the bHLH DNA-binding equipment. Importantly, this approach can be easily adapted for high-throughput studies of structure–function relationships for any protein.

Basic helix–loop–helix transcription factors were first identified by Baltimore and co-workers³ over ten years ago. The conserved amino acids in the bHLH family are present in two α -helices (helix 1 and helix 2, respectively), which are separated by a loop region of variable length and amino-acid

composition^{4,5}. The bHLH domain of several transcription factors is necessary and sufficient for the required dimerization of the protein, and subsequent recognition of six specific nucleotides in the target DNA sequence. Residues in the amino terminus of helix 1 provide sequence-specific DNA recognition, whereas dimerization is largely mediated by helix 2 and the carboxy terminus of helix 1 (Fig. 1, overleaf).

What Winston and Gottesfeld² have done is to use a combinatorial strategy to look at the role of the loop region in a *Drosophila* bHLH factor, Deadpan, in sequence-specific DNA recognition. Using solid-phase peptide synthesis methods and high-resolution mass spectrometry, they have determined the minimal loop requirements necessary to retain full DNA-binding activity. Furthermore, they have identified a loop residue that has a key part in binding DNA through elements that lie outside the core-DNA recognition sequence.

The studies of Winston and Gottesfeld involved parallel approaches to generate

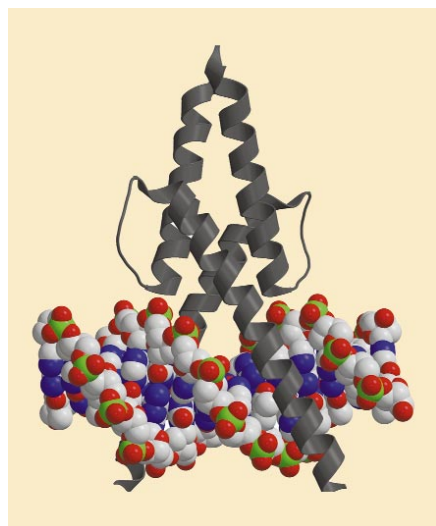


Figure 1 Three-dimensional representation of the structural elements of a basic helix-loop-helix (bHLH) protein (grey). Note that sequence-specific DNA recognition requires the formation of a bHLH protein dimer. Winston and Gottesfeld² show that contacts between residues in the loop regions and base pairs outside the core recognition sequence contribute significantly to energetically favourable DNA-protein interactions. This figure is based on the co-crystal structure of the bHLH protein MyoD bound to its DNA recognition element⁵.

different libraries of Deadpan variants — one in which the length of the loop region was modulated, and another in which the composition of the 12 amino acids that make up the full-length loop was systematically varied. Using stepwise solid-phase peptide synthesis, the first bHLH library, containing successive single amino-acid deletions from both the amino- and the carboxy-terminal ends of the loop region, was quickly generated. This approach greatly reduces the drudgery associated with traditional recombinant DNA techniques, such as gene construction, protein expression and purification, and then characterization of the individual gene products. Twenty-six bHLH loop variants were generated in a few days, the time usually required to produce a single construct by recombinant techniques.

Winston and Gottesfeld took advantage of the power of solid-phase peptide synthesis to produce a second library of bHLH variants containing unnatural peptidyl linkers. Specifically, an alanine-*O*-glycine depsipeptide unit (a peptide that contains an amide-to-ester substitution) was systematically scanned through 11 positions in the Deadpan loop. Substitution of the depsipeptide unit allows removal of the side chains of adjacent amino acids (the chemical equivalent of alanine scanning). The positions of the deleted side chain can be readily identified by selective cleavage of the ester linkage

followed by characterization of the resultant cleavage products by matrix-assisted laser desorption mass spectrometry.

The DNA-binding activities of both libraries were quickly screened by using a sequence-specific DNA-affinity resin containing the DNA recognition element. Deletion variants of the loop region that retained the ability to bind the resin with normal affinity were characterized in a high-throughput fashion with the mass spectrometry technique. Winston and Gottesfeld found that Deadpan tolerates only a small deletion of the centre of the loop without severely compromising DNA binding.

The identity of amino acids in the loop that might contribute to DNA binding was similarly assessed by DNA selection, followed by selective cleavage coupled with mass spectrometry. The authors show that the replacement of a single lysine residue (lysine 80 in Deadpan) by the depsipeptide unit severely reduces DNA-binding affinity. They further demonstrate that replacement of lysine 80 by unnatural amino acids greatly affects DNA affinity, implicating this loop residue as a central player in DNA binding. In the co-crystal structure of a related bHLH factor, Max, the corresponding lysine (lysine 57) makes contacts with the DNA phosphate backbone, and does so in a region outside the core-DNA recognition element.

This work tells us a great deal about sequence-specific DNA recognition, especially as it pertains to a related bHLH family

member, the proto-oncoprotein Myc. The search for Myc target genes has often been based solely on the presence of a six-nucleotide, core Myc-binding site in the promoter region of these target genes. But there are two problems with this strategy. First, the Myc recognition sequence is also recognized by several other members of the bHLH class (including Deadpan); second, Myc binding still occurs *in vitro* even if the base pairs within the core are changed⁶.

Evidence has been accumulating that nucleotides outside the core DNA sequence might influence binding; indeed, detailed examination of Myc-binding sites *in vivo* has revealed a predominance of guanine-cytosine dinucleotides flanking the core recognition sequence⁷. If, as Winston and Gottesfeld show, interactions between bHLH loop residues and DNA occur outside the core six nucleotides, it seems plausible that the effects of flanking nucleotides might further restrict the number of possible Myc-binding sites within the genome. Moreover, given that the largest variations in sequence between members of the bHLH family occur in the loop regions¹, the results of Winston and Gottesfeld also pave the way for identifying a molecular basis for how different bHLH proteins discriminate between all possible binding sites.

More generally, the new work shows the power of chemical synthesis methods coupled with high-resolution mass spectrometry. Winston and Gottesfeld screened and

Astronomy

The stellar fuel

This false-colour image of the nearby galaxy IC342 shows the relationship between gas, stars and star formation. The blue features are clusters of stars, whereas the red in the outer region is the atomic hydrogen gas. Green indicates the molecular hydrogen gas, out of which new stars are forming, and yellow regions are where the atomic and molecular gas overlap.

This map of the molecular gas was obtained by Jean Turner of the University of California at Los Angeles, using 'on the fly' mapping at the 12-metre telescope of the National Radio Astronomy Observatory at Kitt Peak.

IC342 is a spiral galaxy, like the Milky Way, and the spiral structure is obvious in the distribution of the gas, which changes from



predominantly molecular in the inner region, to predominantly atomic in the outer part. The gas forms a bar-like structure around the nucleus, which is believed to play a role in funnelling gas towards the centre, thereby supplying the fuel to form new stars. Curiously, the green ring around the nucleus, which indicates a lot of molecular gas, does not seem to be associated with the formation

of new stars.

The relationship between the phases of gas and the formation of new stars is still not completely understood. Although it is apparent that in many places simply having lots of gas on hand is sufficient to produce stars, this is not universally true. As we produce higher-resolution maps of the distribution of gas in galaxies, we will eventually sort out the physics of what is really happening.

Leslie Sage

fully characterized all possible Deadpan loop deletion and substitution variants (some 42 proteins in all) without purifying a single protein. Furthermore, the use of chemical synthetic methods permits the incorporation of amino acids not found in nature. As all of the steps involved in combinatorial synthesis and characterization can be automated, high-throughput functional screening of any synthetic protein domain is within reach. The availability of gene products from any open reading frame from the many sequenced genomes has spawned high-throughput crystallography for identifying new structural domains; the techniques used by Winston and Gottesfeld give us a glimpse of ways to characterize these new domains functionally as well. The

era of functional genomics is upon us. ■

Satish K. Nair and Stephen K. Burley are in the Laboratories of Molecular Biophysics and the Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

e-mail: nairs@rockvax.rockefeller.edu

1. Littlewood, T. & Evan, G. I. *Helix-Loop-Helix Transcription Factors* (Oxford Univ. Press, New York, 1998).
2. Winston, R. L. & Gottesfeld, J. M. *Chem. Biol.* **7**, 245–251 (2000).
3. Murre, C., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777–783 (1989).
4. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. *Nature* **363**, 38–45 (1993).
5. Ma, P. C., Rould, M. A., Weintraub, H. & Pabo, C. O. *Cell* **77**, 451–459 (1994).
6. Grandori, C. & Eisenman, R. N. *Trends Biochem. Sci.* **22**, 177–181 (1997).
7. Grandori, C., Mac, J., Siebelt, F., Ayer, D. E. & Eisenman, R. N. *EMBO J.* **15**, 4344–4357 (1996).

Non-equilibrium physics

Freezing by heating

H. Eugene Stanley

Everyone has noticed that cooperation among individuals can become less efficient when the individuals are nervous or lack a clear direction. Why is this? Writing in *Physical Review Letters*, Dirk Helbing, Illés Farkas and Tamás Vicsek¹ show that increasing the erratic motions of interacting entities can lead to the breakdown of an efficient pattern of interactions and finally produce a lasting deadlock. This work may be relevant to the movement of light (rising) and heavy (sinking) particles in a vertical column of fluid, and in a wider context to the movement of pedestrians or vehicles in constrained spaces and under heavy traffic conditions.

Transitions induced by random fluctuations or noise have always fascinated physicists^{2–5}. For example, stochastic resonance is a noise-related effect whereby fluctuations can vastly improve the detection of a signal. Helbing *et al.*¹ study a system consisting of a relatively small number of entities, about 100, driven in opposite directions under the influence of noise and interacting through a simple, repulsive force. Such driven interactions take place under non-equilibrium conditions, which means they do not have to follow the expected behaviour patterns of classical equilibrium systems^{6,7}. In their model, Helbing *et al.* discover a new transition that they call 'freezing by heating'. To order a system by heating violates the thermodynamic principles that govern macroscopic equilibrium systems.

Helbing *et al.*¹ report a transition from a fluid state to a frozen state on increasing the noise amplitude — effectively the 'temperature' of the system (Fig. 1). Whereas the fluid state consists of self-organized lanes

of uniform directions of motion (Fig. 1a), which can be disturbed by sufficiently strong fluctuations, the frozen state corresponds to a crystallized configuration of the entities (Fig. 1c). This crystallized state can be destroyed by extreme fluctuations, giving rise to a third, disordered (or 'gaseous') state with randomly distributed entities. Thus, with increasing 'temperature', the authors observe an atypical sequence of transitions from fluid to solid to gas.

Freezing by heating is the opposite of melting, in which increasing the temperature increases the energy, and order is destroyed. It is also different from noise-induced ordering in glasses or granular media. In that case, increasing the 'temperature' drives the system from a disordered metastable state (corresponding to a local energy minimum) to an ordered stable state (corresponding to the global energy minimum). Instead, freezing by heating shows an increase in order with increasing temperature, although the total energy increases at the same time. Because the entropy (amount of disorder) in the ordered crystallized state is smaller than in the less ordered fluid state, the thermodynamic principle of maximization of free energy does not apply. For similar reasons, the apparently related ordering phenomenon seen in experiments with colloidal particles in confined spaces⁸ does not correspond to freezing by heating, but future experiments may reveal their exact relationship.

Such unusual behaviour is possible only because of the far-from-equilibrium nature of the transition. The organized fluid state (Fig. 1a), characterized by lanes, is usually more stable than the crystallized structure,

because the total energy is lower in this state. The crystallized state is therefore metastable — it is sensitive to structural perturbations, such as the interchange of a few entities. If the interface is rough, as in most spontaneously formed jammed states (Fig. 1b), some entities may break through and form 'channels' so the jam eventually collapses. But if the interface between oppositely moving entities is, by chance, flat enough, a stationary crystal is formed. This ordered state, created with the help of large fluctuations, remains frozen even when the noise is reduced again (Fig. 1c). According to Helbing *et al.*, the formation of ordered states requires both frictional and driving forces acting on each entity. Without friction, both lanes and the crystallized state will eventually give way to a disordered 'gaseous' state.

Consideration of self-driving forces in simple systems of interacting entities has opened a fascinating new area of research⁶. Such work is relevant to understanding the flocking of birds, which can be viewed as a non-equilibrium version of the conventional equilibrium transition to ferromagnetism in systems of interacting spins. It also applies to the collective patterns of motion formed

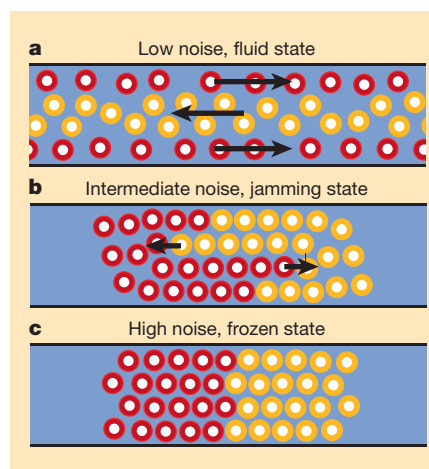


Figure 1 Freezing by heating according to a study by Helbing *et al.*¹. a, For small fluctuation (noise) levels, the oppositely moving entities separate into freely moving lanes. Such a pattern minimizes the interactions between entities, if their density is not too high, and is stable with respect to small disturbances. b, When the noise (or 'temperature') is increased beyond a critical intensity, the fluctuations prevent lane formation. As a consequence, some of the oppositely moving entities block one another locally from time to time, which usually leads to unstable jams with rough interfaces, as new entities arrive at the boundaries of the blocked area. c, If, by chance, the jam develops a 'flat' interface perpendicular to the desired direction of motion and it exists long enough, all the entities arrange themselves in a hexagonal 'crystal' structure in which the driving forces are balanced.

by pedestrians, or to transitions to inefficient, congested traffic states, which can be triggered by fluctuations in the traffic flow. The latter resembles freezing by heating because there is a metastable regime of traffic flow at medium vehicle densities, in which flow is stable under small fluctuations (giving rise to free flow), whereas the build-up of traffic jams (akin to a frozen state) occurs with larger fluctuations. These ideas are essential when trying to prevent traffic breakdowns by suppressing fluctuations in the flow. Industrial estimates of the potential market for traffic-controlling measures, such as intelligent speed limits, on-ramp controls or driver-assistance systems, exceed US\$1 billion a year.

Freezing by heating is one of the most intriguing phenomena found in self-driven systems. A deeper understanding of the role of confined spaces and fluctuations in the transport and jamming of entities is likely to be helpful in developing more efficient designs for granular hoppers and pedestrian facilities. In particular, the freezing-by-

heating transition may apply to the behaviour of pedestrians in panic situations, such as in a smoke-filled room, in which jammed states tend to build up. So perhaps this remarkable phenomenon will interest builders of emergency exits and escape routes⁹. ■
H. Eugene Stanley is at the Center for Polymer Studies, Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA.
e-mail: hes@bu.edu

1. Helbing, D., Farkas, I. J. & Vicsek, T. *Phys. Rev. Lett.* **84**, 1240–1243 (2000).
2. Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. *Rev. Mod. Phys.* **70**, 223–288 (1998).
3. Reimann, P., Bartussek, R., Häußler, R. & Hänggi, P. *Phys. Lett. A* **215**, 26–31 (1996).
4. Nicolis, G. & Prigogine, I. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (Wiley, New York, 1977).
5. Horsthemke, W. & Lefever, R. *Noise-Induced Transitions* (Springer, Berlin, 1984).
6. Helbing, D., Herrmann, H. J., Schreckenberg, M. & Wolf, D. E. *Traffic and Granular Flow '99* (Springer, Berlin, 2000).
7. Laughlin, R. B. *et al. Proc. Natl Acad. Sci. USA* **97**, 32–37 (2000).
8. Rudhardt, D., Bechinger, C. & Leiderer, P. *Phys. Rev. Lett.* **81**, 1330–1333 (1998).
9. Wei, Q.-H., Bechinger, C., Rudhardt, D. & Leiderer, P. *Phys. Rev. Lett.* **81**, 2606–2609 (1998).

Immunology

Commit ye helpers

Anne O'Garra

Different types of immune cell defend the body against different pathogens. For example, immune cells known as type 1 helper T cells (T_H1 cells) are important for protection against intracellular pathogens such as bacteria, and are also implicated in organ-specific autoimmune diseases^{1–5}. T_H2 cells, by contrast, are needed to eradicate flatworms, roundworms and other extracellular parasites, and are also involved in allergic reactions^{1–4}. Part of what makes T_H1 and T_H2 cells different is that they express different types of secreted proteins known as cytokines^{1–5}, and, conversely, different cytokines also have a say in whether a specific progenitor cell becomes a T_H1 or a T_H2 cell^{3,4}. But other mechanisms are also likely to be involved in the commitment of a progenitor cell to a particular differentiation pathway^{4–6}.

Now, Szabo and colleagues⁷, writing in *Cell*, identify such a mechanism. They report the isolation of T-bet, a new T_H1 -specific factor that, remarkably, not only controls the expression of interferon- γ — the hallmark T_H1 cytokine — but also represses T_H2 -specific cytokines.

What did we already know about the differentiation of T-helper cells? Have a look at Fig. 1 (overleaf). A particular type of naive T cell can develop into a T_H1 or a T_H2 cell. Development of T_H1 cells is driven by a cytokine called interleukin-12, which is

produced by immune cells known as macrophages and dendritic cells. Interleukin-12 induces the naive T cell to produce the signature T_H1 -cell cytokine, interferon- γ . Commitment to the T_H1 lineage is, in turn, enhanced by interferon- γ , which upregulates expression of the interleukin-12 receptor while inhibiting the growth of T_H2 cells^{4,5}. Another protein that is important in T_H1 differentiation is Stat4 (for signal transducer and activator of transcription-4), a transcription factor found in the cell's cytoplasm⁵.

On the other hand, the cytokine interleukin-4 induces T_H2 development and the production of interleukins 4, 5 and 13, through activation of the transcription factor Stat6 (refs 3–5). Interleukin-4 also downregulates expression of the interleukin-12 receptor on developing cells, helping them to commit to the T_H2 lineage^{3–5}. So, a specific genetic programme results in differentiation towards a particular T-helper lineage.

As well as Stat4 and Stat6, other lineage-specific transcription factors are needed for the differentiation of T-helper cells. So far, we know of two more T_H2 -specific transcription factors — c-Maf and GATA-3 (refs 8–10). c-Maf increases the expression of interleukin-4 (ref. 8), while GATA-3 can induce the expression of a broad spectrum of T_H2 -specific cytokines in developing^{9–14} and committed¹⁴ T_H1 cells, and inhibit the production of interferon- γ ^{11–14}.

In contrast, we are still lacking crucial details about the molecular basis of T_H1 differentiation. We do know that a protein called ERM is induced by interleukin-12 in a Stat4-dependent manner and is specific to T_H1 cells, but this protein does not affect the production of T_H1 -specific cytokines¹⁵. Now, however, Szabo *et al.*⁷ have filled in the gaps by identifying a previously unknown protein, T-bet, as being key to T_H1 differentiation.

T-bet belongs to the so-called T-box family of transcription factors. These proteins are important in a myriad of developmental processes, but have never cropped up in the immune system before. Szabo *et al.* found that T-bet is rapidly and selectively induced in developing T_H1 , but not T_H2 , cells. When the authors introduced T-bet DNA into a mouse T-cell line, they found significant activation of an interferon- γ gene construct that was introduced at the same time. In contrast, T-bet expression repressed the activation of the interleukin-2 gene, in keeping with previous observations that, although T_H1 cells express interleukin-2 to start with, they downregulate its expression as they differentiate. However, T-bet had no effect on transactivation of the interleukin-4 promoter.

So, T-bet induces expression of the key T_H1 cytokine, interferon- γ , in this T-cell line. Interestingly, Szabo *et al.* also showed that mixed populations of B cells and natural killer cells (yet more immune cells) showed increased T-bet expression after being cultured with, among other proteins, interleukin-12. Only small amounts of interferon- γ appeared in these cells, but these levels increased after addition of interleukin-18, as did the expression of T-bet. Perhaps interleukin-18 induces another factor required for interferon- γ production⁴, or perhaps T-bet must be expressed above a certain threshold level to achieve significant induction of interferon- γ ⁷. Indeed, other transcription factors affect the determination of other lineages in a dose-dependent way¹⁶. In any case, it seems that interleukin-12 induces expression of T-bet, and that this is enhanced by interleukin-18. In turn, T-bet, possibly with other factors, increases expression of interferon- γ . This theory is supported by other recent findings¹⁷.

Szabo *et al.*⁷ also found that T-bet diverts naive T cells into the T_H1 pathway, and, remarkably, can even convert committed T_H2 cells into T_H1 cells. First, the authors expressed T-bet in T cells that had been cultured for a short time together with stimuli important for T_H2 differentiation. Expression of T-bet induced a large number of these cells to produce interferon- γ , and led to a marked reduction in the number of interleukin-4-producing cells, independently of interferon- γ .

Stable commitment to a particular helper

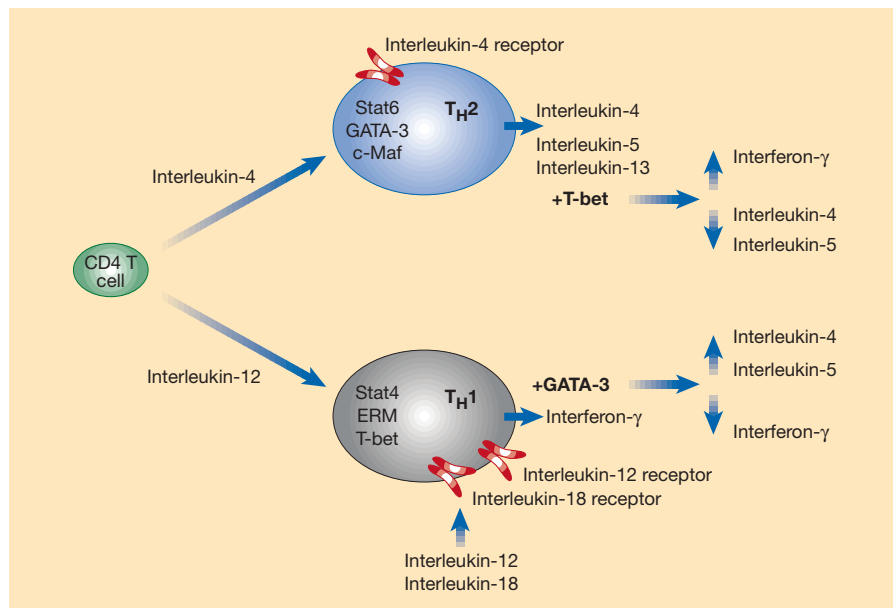


Figure 1 Driving T-helper-cell commitment. The cytokines interleukin-4 and interleukin-12 direct the development of, respectively, T-helper-2 (T_H2) and T_H1 subsets from a naive T cell. (Expression of the protein CD4 defines a subset of T cells, including the naive T-helper cell.) Acting through the transcription factors Stat6 and Stat4, these cytokines lead T_H2 cells to produce interleukins 4, 5 and 13, and T_H1 cells to produce interferon- γ . T_H2 cells express the transcription factors c-Maf and GATA-3. c-Maf activates the expression of the interleukin-4 gene. Expression of GATA-3 in developing and committed T_H1 cells induces the expression of a wide array of T_H2 -specific cytokines and inhibits interferon- γ production. Szabo *et al.*⁷ have now identified T-bet, a T_H1 -specific transcription factor that is induced by interleukin-12. T-bet itself induces expression of interferon- γ and represses expression of interleukins 4 and 5 in both developing and committed T_H2 cells.

subset occurs after repeated stimulation under T_H1 -specific or T_H2 -specific culture conditions¹⁸. Szabo *et al.* next introduced T-bet into such stably committed T_H2 populations. They found that the number of cells producing interferon- γ increased, and the number of interleukin-4- and interleukin-5-producing cells decreased (Fig. 1). It seems that T-bet can convert even committed T_H2 into T_H1 cells.

However, introduction of T-bet into a more long-term, unvarying T_H2 clone increased the number of interferon- γ -producing cells only slightly, perhaps because T-bet could no longer access the relevant genes, or because the array of factors necessary for inducing expression of the interferon- γ gene was no longer present. Yet T-bet expression in this clone still resulted in a significant reduction in the number of interleukin-4- and interleukin-5-producing cells, so T-bet's ability to suppress T_H2 development extends even to long-term clones.

All of this means that T-bet both initiates T_H1 developmental programmes and represses the opposing T_H2 programmes. This effect of T-bet ties in nicely with the ability of GATA-3 to repress interferon- γ in T_H1 cells^{11–14} while simultaneously inducing T_H2 -specific cytokines^{9–14}. Also, like T-bet, GATA-3 affects the differentiation of even committed T-helper cells¹⁴ (Fig. 1). So, the balance of T-bet and GATA-3 ultimately

determines the fate of a developing T-helper cell. These findings should lead to further insight into the molecular mechanisms that direct lineage commitment in the immune system, and may also provide scope for therapeutic intervention. In the meantime, we need to know what determines the relative prominence of these two important T-helper aides.

Anne O'Garra is at the DNAX Research Institute (a Schering Plough Institute), Palo Alto, California 94304-1104, USA.

e-mail: aog@dnax.org

1. Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. *J. Immunol.* **136**, 2348–2357 (1986).
2. Sher, A. & Coffman, R. L. *Annu. Rev. Immunol.* **10**, 385–409 (1992).
3. Abbas, A. K., Murphy, K. M. & Sher, A. *Nature* **383**, 787–793 (1996).
4. O'Garra, A. *Immunity* **8**, 275–283 (1998).
5. Murphy, K. M. *et al. Curr. Top. Microbiol. Immunol.* **238**, 13–26 (1999).
6. Agarwal, S. & Rao, A. *Immunity* **9**, 765–775 (1998).
7. Szabo, S. J. *et al. Cell* **100**, 655–669 (2000).
8. Ho, I.-C., Hodge, M. R., Rooney, J. W. & Glimcher, L. H. *Cell* **85**, 973–983 (1996).
9. Zheng, W.-P. & Flavell, R. A. *Cell* **89**, 587–596 (1997).
10. Zhang, D. H., Cohn, L., Ray, P., Bottomly, K. & Ray, A. *J. Biol. Chem.* **272**, 21597–21603 (1997).
11. Ouyang, W. *et al. Immunity* **9**, 745–755 (1998).
12. Ouyang, W. *et al. Immunity* **12**, 27–37 (2000).
13. Ferber, I. A. *et al. Clin. Immunol.* **91**, 134–144 (1999).
14. Lee, H. J. *et al. J. Exp. Med.* (in the press).
15. Ouyang, W. *et al. Proc. Natl Acad. Sci. USA* **96**, 3888–3893 (1999).
16. McDevitt, M. A., Shivdasani, Y., Fujiwara, H., Yang, H. & Orkin, S. H. *Proc. Natl Acad. Sci. USA* **94**, 6781–6785 (1997).
17. Heath, V. L. *et al. J. Immunol.* **164**, 2861–2865 (2000).
18. Murphy, E. *et al. J. Exp. Med.* **183**, 901–913 (1996).

Daedalus

Rain, rain, go away

The cork in a wine bottle, says Daedalus, is an extremely cunning invention. It exploits the simple fact that cork swells in high humidity. The inner end of the cork, sealing the airspace over the wine, is at well over 90% humidity. So once in place, the cork swells and seals the bottle perfectly.

Daedalus is now applying this elegant principle to paint. In rainy Britain, every exposed wooden surface must be painted to stop it rotting. But no paint is totally impermeable to water or its vapour. It merely slows the rate at which the wood takes up or loses water. This is useful because even in Britain it rains only 5% of the time. A coat of paint prevents the wood imbibing a great draught of water every time it rains, and keeps its water content in balance with the average humidity. This is usually low enough to discourage rot.

But Daedalus is going beyond such purely passive barrier films. His new paint will let water out much more easily than it gets in. It will be inherently porous; but, like cork, it will swell in high humidity, sealing its pores against the ingress of water. Conversely, in hot or dry weather it will shrink, and its pores will open to let water vapour out easily. It will be an active water-pump, keeping a wooden surface drier on average than the air around it.

DREADCO's 'Pumpaint' will combine air-polymerization and emulsion-paint technology. It will consist essentially of a suitable monomer dissolved or suspended in excess water. When the paint is applied to a wooden surface, its monomer component will set quickly. As its water later evaporates, the polymer will dry out, shrinking into a microporous film. In humid conditions, it will take up water vapour again and re-expand to a homogeneous barrier. This constant stretching and shrinking risks peeling it from the wood beneath. It will have to bind itself firmly by chemical reaction.

Pumpaint will lay to rest the humidity paranoia of British builders, joiners and carpenters. They will enjoy once more the craftsmanship and feel of wood, and the elegance of its products. The dour advance of PVC, whose only advantage is that it does not rot, will be halted. And many noble ancient houses, steadily sabotaged by their deteriorating woodwork, may yet be saved.

David Jones

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special Nature offer: m.curtis@nature.com

Delaying the onset of Huntington's in mice

This unremitting disease develops later in animals stimulated by their environment.

Huntington's disease is an inherited (autosomal dominant) disorder in which there is progressive neurodegeneration, affecting the corpus striatum and cerebral cortex of the brain, and for which there is no known cure. Transgenic mice have been created^{1,2} that develop a neurodegenerative syndrome that closely models the human disease. Here we show that exposure of these mice to a stimulating, enriched environment from an early age helps to prevent the loss of cerebral volume and delays the onset of motor disorders.

We randomly allocated 30 male Huntington's disease (HD) R6/1 mice to either a normal or a stimulating environment. Sixteen of these mice were positive for the HD transgene, in which exon 1 of the human *huntingtin* gene¹ is expressed with an expanded CAG repeat² encoding an extended polyglutamine tract. All mice were housed in groups in large standard cages. Routine care included provision of normal feed and bedding, whereas the cages of the 'environmentally enriched' groups also contained cardboard, paper and plastic objects, which were changed every two days, from the age of 4 weeks.

We characterized the response of the mice to their new environment by testing each mouse once, at 19–22 weeks of age, in an experimental cage, where it had the option of approaching or avoiding such stimuli. The cage was identical to the home cages except that only one half contained 'enrichment' objects. The mouse was put in the middle of the cage and the time spent in each half of the cage was measured over 3 min. Mice from the enriched groups spent much of their time exploring the objects (75% for wild-type and 70% for HD mice).

Environmental enrichment was not associated with any effect on body mass or spontaneous motor activity. HD mice were on average 16% lighter at 20 weeks than wild-type littermates (29.6 g versus 35.3 g; $P < 0.01$), but enrichment had no significant effect on body weight in either group. Blood and urine glucose levels were normal in all mice.

To define the onset of disease, motor coordination was tested every week in a 'turning task', by placing each mouse at the end of a suspended, horizontal wooden rod. This test reveals deficits in motor coordination while minimizing the influence of memory and muscular strength. Strikingly, only one of the environmentally enriched group of HD mice (14%) had developed this behavioural sign at the end of testing at 22 weeks (Fig. 1a; $P < 0.000001$).

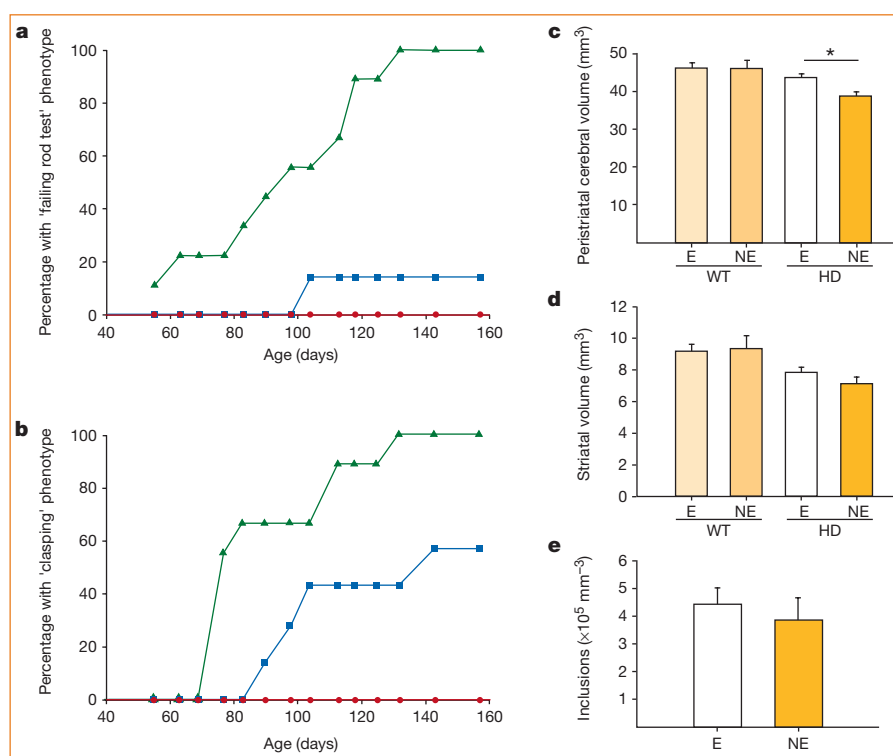


Figure 1 Exposure to a stimulating environment delays the onset of the disease phenotype in R6/1 HD mice. **a**, The cumulative percentage of mice consistently failing a test of motor coordination is plotted as a function of age. A failure is defined as consistent falling or inability to turn around on a suspended horizontal rod. The enriched environment markedly delayed the appearance of this behavioural sign in HD mice ($P < 0.000001$; χ^2 test). **b**, This environment also delayed the appearance of sustained rear-paw claspings, another indication of HD phenotype in R6/1 mice. The cumulative percentage of mice reaching the onset of this sign is plotted as a function of age. Onset was highly significantly delayed in the enriched environment ($P < 0.000001$; χ^2 test). Squares, HD mice in enriched environment; triangles, HD mice given routine care; circles, wild-type (WT) mice (enriched (E) and non-enriched (NE) groups pools: no failures). **c**, Effect of environmental enrichment on the volume of the corpus striatum and the surrounding portion of the entire forebrain in HD mice. The outlines of the whole brain and of the striatum on both sides were digitized for a series of coronal sections running from the caudal to the rostral pole of the striatum. The 'peristriatal cerebral volume' was calculated separately for E and NE groups of WT and HD mice. Analysis of variance (ANOVA) indicates a significant effect of enrichment ($F[1,21] = 4.63$, $P = 0.043$), a highly significant effect of genotype ($F[1,21] = 16.17$, $P = 0.0006$) and an enrichment $\times$ genotype interaction that did not reach significance ($F[1,21] = 4.18$, $P = 0.054$). Asterisk, $P < 0.01$, Student's *t*-test. **d**, ANOVA of differences between the groups in the volume of the striatum (both sides of the brain combined) indicates no effect of enrichment ($F[1,21] = 0.394$, $P = 0.54$), a significant effect of genotype ($F[1,21] = 18.89$, $P = 0.0003$) and no enrichment $\times$ genotype interaction ($F[1,21] = 1.00$, $P = 0.33$). **e**, Exposure to an enriched environment does not retard the appearance of ubiquitinated neuronal inclusions in the corpus striatum of R6/1 HD mice. The densities of ubiquitinated inclusions (means $\pm$ s.e.m.) were estimated from counts through the entire striatum of E and NE HD mice. They did not differ significantly ($P = 0.55$, Student's *t*-test).

Another early sign of disease in HD mice is claspings of the rear paws when briefly suspended by the tail². This sign is presumably independent of muscle strength and acquired motor skills. Again, the appearance of this sign was significantly delayed in the enriched group compared with the normally reared HD animals (Fig. 1b; $P < 0.000001$).

A late component of the disease phenotype in HD mice is the onset of seizures². Two of the non-enriched group had seizures, at 14 and 19 weeks old respectively, but none of the enriched mice had them.

As disease was advancing in the control HD mice, we examined their brains at 22 weeks of age by quantitative histology using

an automated stereological analysis system (Neurolucida).

The combined volume of the striatum, 'striatal volume', on both sides of the brain was calculated, as was the volume of the entire coronal block of cerebrum between the rostral and the caudal poles of the striatum, which includes the part of the cerebral cortex with the densest projections to the striatum. This 'peristriatal cerebral volume' was 13% larger in the environmentally enriched HD mice than in the non-enriched HD group (Fig. 1c; $P < 0.05$), whereas enrichment did not affect the peristriatal cerebral volume of wild-type mice. The striatum was slightly larger in enriched than

in non-enriched HD mice, but this difference was not significant (Fig. 1d; $P=0.33$). These results suggest that environmental enrichment delays the degenerative loss of cerebral volume in HD mice and that the corticostriatal pathway plays a role in the pathogenesis of HD.

HD mice develop neuronal inclusions that are immunopositive for ubiquitin³. There was no significant difference in the overall density of inclusions³ in the striatum between enriched and non-enriched HD mice (Fig. 1e). This shows that enrichment does not prevent the formation of inclusions, despite its effects on the appearance of symptoms. The TUNEL assay revealed no occurrence of apoptotic cell death.

The remarkable effect of environmental enrichment in delaying the onset of neurological signs in HD mice could be explained by increased sensory input and/or motor activity having a direct influence on the striatum, for example by altering the synaptic number and morphology of medium spiny neurons^{4,5}. Alternatively, the effect could be mediated through the cortical input. Environmental enrichment might, for instance, counter the reduction in activity in parts of the cerebral cortex that occurs in Huntington's disease⁶.

Deficits in a form of synaptic plasticity — long-term potentiation (LTP) — have been found in hippocampal slices from HD mice^{7,8}. Environmental enrichment in rats increases the strength of specific hippocampal synapses, regulates the induction of LTP, increases the binding of glutamate to the AMPA receptor⁹, and protects against excitotoxicity¹⁰. Environmental enrichment may therefore overcome deficiencies of synaptic plasticity, particularly at intracortical and corticostriatal^{4,5} synapses, and help to improve the defects in HD mice.

Monozygotic twins with Huntington's disease and identical CAG-repeat lengths can display distinctly different clinical symptoms and behavioural abilities, implying that the disease is affected by environmental factors¹¹. Our findings suggest that occupational therapy based on the principles of environmental enrichment might delay the onset of Huntington's disease in humans as well.

Anton van Dellen*, Colin Blakemore*, Robert Deacon†, Denis York‡, Anthony J. Hannan*

*University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK

†Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK

‡Department of Virology, University of Natal Medical School, Durban, South Africa
e-mail: avd@physiol.ox.ac.uk

1. Huntington's Disease Collaborative Research Group *Cell* **72**, 971–983 (1993).

2. Mangiarini, L. *et al.* *Cell* **87**, 493–506 (1996).

3. Davies, S. W. *et al.* *Cell* **90**, 537–548 (1997).

4. Comery, T. A., Shah, R. & Greenough, W. T. *Neurobiol. Learn. Mem.* **63**, 217–219 (1995).

5. Comery, T. A., Stamoudis, C. X., Irwin, S. A. & Greenough, W. T. *Neurobiol. Learn. Mem.* **66**, 93–96 (1996).

6. Deckel, A. W., Cohen, D. & Duckrow, R. *Neurology* **51**, 1576–1583 (1998).

7. Murphy, K. P. S. J. *et al.* *J. Neurosci.* (in the press).

8. Hodgson, J. G. *et al.* *Neuron* **23**, 181–192 (1999).

9. Foster, T. C., Gagne, J. & Massicotte, G. *Brain Res.* **736**, 243–250 (1996).

10. Young, D. H., Lawlor, P. A., Leone, P., Dragunow, M. & During, M. J. *Nature Med.* **5**, 448–453 (1999).

11. Georgiou, N. *et al.* *Mov. Disord.* **14**, 320–325 (1999).

Biogeochemistry

Hexadecane decay by methanogenesis

The potential for the biological conversion of long-chain saturated hydrocarbons to methane under anaerobic conditions has been demonstrated by using an enrichment culture of bacteria to degrade pure-phase hexadecane¹. The formation of methane in hydrocarbon-rich subsurface zones could be explained if a similar conversion of long-chain alkanes to methane were to take place in subsurface environments. If this process could be stimulated in the subsurface, it could be used to enhance hydrocarbon recovery from petroleum reserves^{1,2}. Parkes², however, questions the environmental significance of the enrichment-culture results¹ on the grounds that alkane conversion to methane is very slow and because sulphate-reducing and methanogenic bacteria might both be necessary for even this slow process to occur, restricting the conversion to specialized, unusual zones in sediments. Here we show that, on the contrary, subsurface bacteria can adapt to convert hexadecane to methane rapidly and in the absence of sulphate-reducing bacteria.

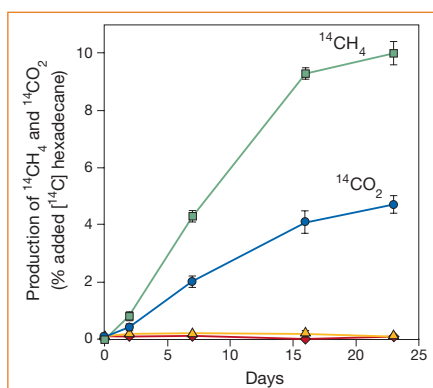


Figure 1 Metabolism of [¹⁴C]hexadecane in contaminated aquifer sediments collected from a subsurface crude-oil spill site near Bemidji, Minnesota. Methanogenic oil-containing sediments readily produced ¹⁴CH₄ (squares) and ¹⁴CO₂ (circles), whereas there was no metabolism of hexadecane in oil-free, anaerobic sediments (¹⁴CH₄, diamonds; ¹⁴CO₂, triangles) collected further downgradient within the contaminant plume. Results are the means of triplicate incubations; error bars show the standard error of the means.

The study of microbial metabolism in deep petroleum reservoirs is problematic owing to the technical difficulty of sampling sediments from these environments. Shallow petroleum-contaminated subsurface environments provide a likely analogue for petroleum reservoirs because microbial processes found in shallow subsurface environments are commonly found in deeper subsurface environments as well³.

We sampled oil-bearing sediments from an aquifer located in Bemidji, Minnesota, that had been contaminated with crude oil. The sediments were incubated under strict anaerobic conditions at 20 °C as previously described⁴. The sediments were depleted of nitrate (to less than 5 μM) and sulphate (to less than 10 μM) and more than 95% of the HCl(0.5 M)-extractable iron was in the Fe(II) state. These sediments actively produced methane over time. When [2-¹⁴C]acetate was injected into the sediments, ¹⁴CH₄ and ¹⁴CO₂ were produced in a ratio of 4:1, which is typical in sulphate-depleted sediments where methane production is the electron-accepting process. The concentration of hydrogen was 5–8 nM, characteristic of sulphate-depleted, methanogenic sediments³.

When we injected [¹⁴C]hexadecane into the sediments, it was converted to ¹⁴CH₄ without a lag (Fig. 1). This demonstrated that the microorganisms were already adapted for conversion of hexadecane to methane and suggests that hexadecane was being degraded *in situ*. In contrast, sediments that were slightly downgradient from the oil zone, and which were also anaerobic but not exposed to alkanes, did not degrade hexadecane. The ratio of ¹⁴CH₄ to ¹⁴CO₂ produced from [¹⁴C]hexadecane in the oil-containing sediments was consistent with the ratio expected¹ for hexadecane conversion under methanogenic conditions.

Parkes' reservations², together with reports that alkanes are not significantly converted to methane in actual anaerobic sediments^{5,6}, have raised questions about how well the results from the enrichment culture experiment¹ can be extrapolated to typically sulphate-depleted, methanogenic subsurface environments. Our results show that subsurface microbial communities can adapt to rapidly convert alkanes to methane in the complete absence of sulphate reduction. The conversion of alkanes to methane is likely to be significant in many petroleum-containing anaerobic environments, including petroleum reservoirs.

Robert T. Anderson, Derek R. Lovley

Department of Microbiology, University of Massachusetts, Amherst, Massachusetts 01003, USA
e-mail: dlovley@microbio.umass.edu

1. Zengler, K., Richnow, H. H., Rossello-Mora, R., Michaelis, W. & Widdel, F. *Nature* **401**, 266–269 (1999).

2. Parkes, J. *Nature* **401**, 217–218 (1999).

3. Lovley, D. R. & Chapelle, F. H. *Rev. Geophys.* **33**, 365–381 (1995).

4. Anderson, R. T., Rooney-Varga, J. N., Gaw, C. V. & Lovley, D. R. *Environ. Sci. Technol.* **32**, 1222–1229 (1998).
5. Ward, D. M., Atlas, R. M., Boehm, P. D. & Calder, J. A. *Ambio* **9**, 277–283 (1980).
6. Ward, D. M. & Brock, T. D. *Geomicrobiol. J.* **1**, 1–9 (1978).

Insect behaviour

Parasitic honeybees get royal treatment

Since the human-assisted movement of the Cape honeybee *Apis mellifera capensis* out of its native territory, its workers have invaded colonies of the African honeybee *A. m. scutellata*. When this happens, their ovaries develop and they begin to reproduce¹, which results in the death of the *scutellata* queen, and eventually in either the death of the colony or the production of a *capensis* (worker-produced) queen¹. We have found that *capensis* larvae alter the behaviour of non-*capensis* workers and receive royal treatment, resulting in adult females with queenlike characteristics (pseudoqueens^{2–4}).

Only twenty years ago, it was feared that *A. m. scutellata* would out-compete *A. m. capensis*², but instead *capensis* workers have invaded *scutellata* colonies⁵. This causes 'dwindling colony' syndrome¹, in which the *scutellata* queen is rejected and *capensis* workers start laying eggs. Unlike the workers of other *Apis* races, which lay haploid eggs, workers of *capensis* produce diploid eggs by thelytoky, which develop into either workers or queens².

The presence of laying *capensis* workers prevents the rearing of an emergency queen⁶. Ultimately, the colony dies, owing to the increase of laying *capensis* workers and a decrease in foraging and brood care; or the laying workers are superseded by a *capensis* queen¹.

Caste in honeybees is normally determined by the selective feeding of female larvae: those destined to be queens are fed royal jelly and receive more food than those destined to be workers, leading to a morphologically distinct queen caste. It is generally assumed that brood in social Hymenoptera possess no power over caste determination, except in social parasites⁷. We found, however, that *capensis* brood can

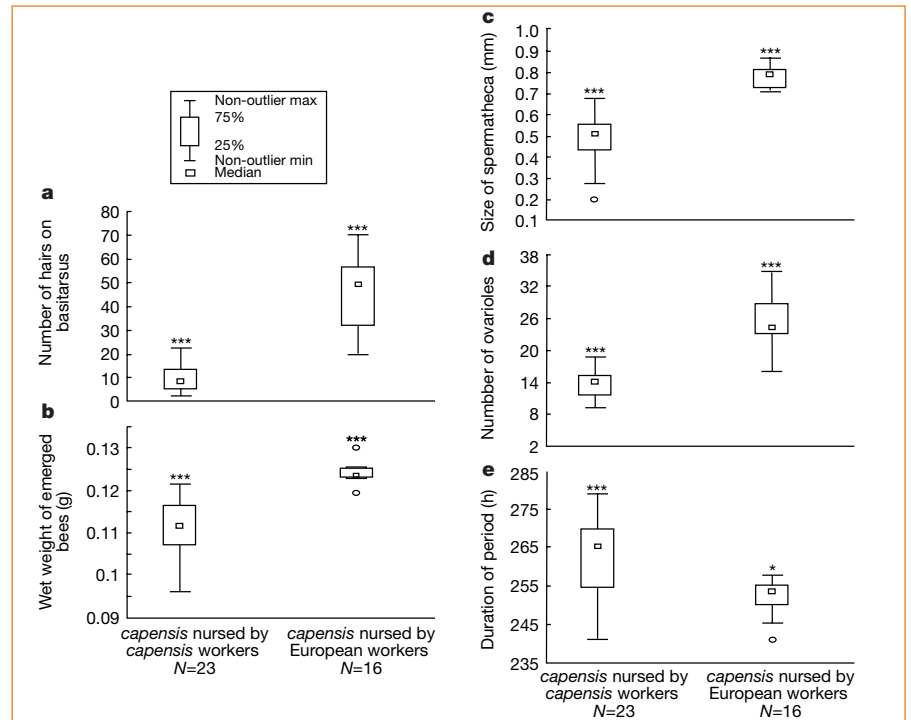


Figure 1 Effect of nursing by European workers. *A. m. capensis* larvae show an increase in the number of hairs between the two lowest pollen combs on the basitarsus (a), fresh weight at emergence (b), size of the spermatheca (c), and total number of ovarioles (d). The duration of the capped period decreased (e). All measured characteristics were significantly different between the two groups (one-way multiple analysis of variance, $P \leq 0.001$ when the whole data set was compared. Characteristics were then tested separately using a post-hoc least-significant difference comparison: * $P < 0.05$, *** $P \leq 0.001$). A frame with *A. m. capensis* eggs (0–24 hours old) was transferred into a European honeybee colony (hybrids of *A. m. carnica* and *A. m. mellifera*). At the same time, a frame of *capensis* eggs laid within the same interval remained in the *capensis* colony. Colonies were of similar size and kept at the same apiary in Wageningen, the Netherlands. Capping of brood cells was registered every three hours. Nine days after capping, combs were placed in an incubator (34.5 °C) and emergence of each specific cell was recorded every 15 min. Fresh weight of bees was measured upon emergence, after which the bees were stored individually in 70% ethanol until further measurements were made. Oval data points are outliers.

control caste fate, which may be important in usurping non-*capensis* colonies.

A. m. capensis brood nursed by European workers are given more food than European larvae (Table 1) — indeed, they receive more food than when they are nursed by *capensis* workers. In addition, the food composition was slightly more similar to royal jelly⁸. This resulted in the production of queen-like *capensis* workers (Fig. 1). The functional worker characteristics (pollen combs and pollen baskets on the hind legs) of *capensis* workers raised in European colonies decreased⁹, whereas their queen characteristics⁹ (short developmental time, weight, a large spermatheca and a large number of ovarioles) increased.

Royal treatment of *capensis* brood by

European workers has the same effect as juvenile hormone⁹. Both affect the complete set of caste characteristics, not just isolated morphological features. Our results show that brood in social Hymenoptera are not necessarily passive: *capensis* brood can manipulate their host so that the resulting workers are more queenlike and are probably better reproducers than normal *capensis* workers. Thus *capensis* is a well adapted social parasite of other bee races.

Madeleine Beekman*†, Johan N. M. Calis*, Willem Jan Boot*

*Laboratory of Entomology, Wageningen Agricultural University, PO Box 8031, 6700 EH Wageningen, The Netherlands

†Present address: Department of Animal and Plant Sciences, Sheffield University, Sheffield S10 2TN, UK

e-mail: M.Beekman@Sheffield.ac.uk

Table 1 Worker feeding of larvae of different origin

	European larvae in European colony	capensis larvae in European colony	capensis larvae in capensis colony
Total dry weight (mg)	10 ± 2 ^a	29 ± 4 ^b	5 ± 1 ^c
Fraction of sugars	0.54 ± 0.07 ^a	0.50 ± 0.05 ^{ab}	0.56 ± 0.03 ^{ac}
Fraction of fructose	0.62 ± 0.04 ^a	0.58 ± 0.02 ^b	0.59 ± 0.06 ^{ab}

Sugar fractions were determined using anthrone reagents¹⁰. *A. m. capensis* larvae reared by European workers received significantly more food and less sugar than simultaneously nursed European workers. When nursed in a European colony, food fed to *capensis* larvae contained less sugar and the fraction of fructose was lower. These qualitative differences influence caste determination². Food fed to workers normally contains more sugar than royal jelly but less fructose⁸ (fraction of sugar and fructose in royal jelly is 0.23 and 0.38, respectively⁸). Food was collected on the fifth day after the eggs were laid (eggs differed by 0–24 hours in age), at which time the cells were nearly capped. Each sample consisted of five brood cells ($n = 10$) and was lyophilized overnight. Values followed by the same letter (a, b, c) were not statistically significant at the $P = 0.05$ level (based on a post-hoc least-significant difference comparison).

1. Hepburn, H. R. & Allsopp, M. H. S. *Afr. Tydsk. Wetenskap* **90**, 247–249 (1994).
2. Ruttner, E. *Apidologie* **8**, 281–294 (1977).
3. Crewe, R. M. & Velthuis, H. H. W. *Naturwissenschaften* **67**, 467–469 (1980).
4. Moritz, R. F. A., Kryger, P. & Allsopp, M. H. *Nature* **384**, 31 (1996).
5. Greeff, J. M. S. *Afr. J. Sci.* **93**, 306–308 (1997).
6. Allsopp, M. S. *Afr. Bee J.* **64**, 52–57 (1992).
7. Nonacs, P. & Tobin, J. E. *Evolution* **46**, 1605–1620 (1992).
8. Brouwers, E. V. M. *J. Apic. Res.* **23**, 94–101 (1984).
9. Wirtz, P. & Beetsma, J. *Ent. Exp. Appl.* **15**, 517–520 (1972).
10. Van Handel, E. *Anal. Biochem.* **19**, 193–194 (1967).

Specificities of heparan sulphate proteoglycans in developmental processes

Norbert Perrimon* & Merton Bernfield†

* Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

† Departments of Pediatrics and Cell Biology, Harvard Medical School, Children's Hospital, Boston, Massachusetts 02115, USA

Heparan sulphate proteoglycans are abundant cell-surface molecules that consist of a protein core to which heparan sulphate glycosaminoglycan chains are attached. The functions of these molecules have remained mostly underappreciated by developmental biologists; however, the actions of important signalling molecules, for example Wnt and Hedgehog, depend on them. To understand both the mechanisms by which ligands involved in development interact with their receptors and how morphogens pattern tissues, biologists need to consider the functions of heparan sulphate proteoglycans in signalling and developmental patterning.

Biochemical studies and cell-culture assays have implicated heparan sulphate proteoglycans (HSPGs) as co-receptors¹ in processes ranging from mechanical support to functions in adhesion, motility, proliferation, differentiation and morphogenesis (see ref. 2 for a review). Enormous structural heterogeneity can be generated through specific heparin sulphate (HS) chain modifications during their biosynthesis, as well as from the diverse nature of their core proteins. Furthermore, biochemical studies have shown how interactions of cell-surface HSPGs can be specific at various levels. Tissue-specific isoforms of the polymerizing and chain-modifying enzymes can produce HS chains with distinct sequences and macroscopic organizations^{3,4}. Various ligands and their receptors can show selectivity in their binding affinities to distinct HS chain structures^{5,6}. These chains can be attached to one or more specific core proteins, each with a distinctive tissue-specific expression pattern and cellular localization. Finally, HSPGs can be selectively shed from the cell surface to yield soluble effectors.

Despite extensive biochemical evidence that HSPGs have specific functions *in vivo*, the roles of these molecules have not been appreciated, especially by developmental biologists. The prevalent view is that these highly abundant, strongly anionic molecules interact non-specifically with secreted growth factors and extracellular matrix (ECM) components, and have a general function in protecting these ligands from proteolysis or sequestering them in the ECM. Indeed, these ligands are frequently purified by chromatography on columns of heparin, the heavily modified HS-related glycosaminoglycan that is pharmaceutically manufactured from mast-cell-rich tissues⁷. This has contributed to the notion that HSPGs lack specificity and simply act as 'extracellular fly paper' that binds and concentrates ligands at the cell surface.

This idea is rapidly changing, however, because of insights from genetic studies in *Drosophila* and mice. Indeed, genetic studies have begun to reveal dedicated functions for HSPGs in specific signalling pathways involved in cell differentiation and morphogenesis. Here we discuss the developments in the field and illustrate the complex and specific roles of HSPGs *in vivo*. From the analysis of mutant phenotypes that are associated with mutations in either the enzymes that synthesize the HS chains linked to the proteoglycan core proteins or the core proteins themselves, it is now apparent that HSPGs are critical in the interactions between specific extracellular ligands and their signal-transducing receptors.

Enzymes involved in HSPG biosynthesis

Precursor HS chains are initially synthesized in the Golgi as non-

sulphated copolymers attached to HSPG core proteins. The chain is initiated at specific Ser-Gly residues within defined amino-acid sequences on the core protein. The number of such chain attachment sites varies with the core protein, but is typically only between two and four. The first step in HS chain biosynthesis is the enzymatic transfer of xylose from the sugar nucleotide to the hydroxyl group of these serine residues (for details, see ref. 5). Each sugar in a common tetrasaccharide linkage region (-GlcA-Gal-Gal-Xyl) is then transferred from the sugar nucleotide by individual Golgi enzymes. A single b-linked GlcNAc residue is added to the tetrasaccharide linker followed by alternating GlcA and GlcNAc residues catalysed by HS polymerase.

Once this chain is assembled, the individual saccharide units undergo a number of sequential modifications by various Golgi enzymes: *N*-deacetylase/*N*-sulphotransferase (NDST), uronosyl C5-epimerase, 2-*O*-sulphotransferase (2-OST), 6-*O*-sulphotransferase (6-OST) and 3-*O*-sulphotransferase (3-OST) (Fig. 1). The enzymes do not modify all the available sugars in the chain, which

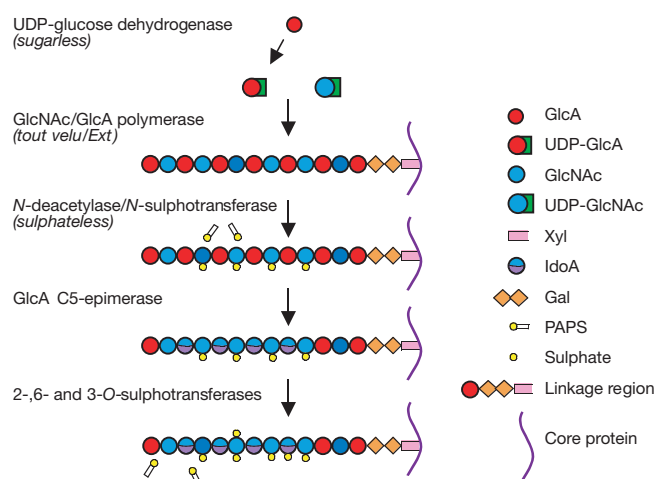


Figure 1 Scheme of HS chain biosynthesis. The chains are synthesized on the core protein by the sequential action of individual glycosyltransferases. A common tetrasaccharide linkage region is formed, followed by the addition of alternating GlcA and GlcNAc residues, producing in turn the precursor chain. This chain is then enzymatically modified by deacetylation and *N*-sulphation, epimerization and *O*-sulphation, yielding individual chains whose sequence is distinct from all other chains.

results in extensive sequence diversity in the final chain. This structural heterogeneity includes a macroscopic organization in which regions of 10–16 highly modified disaccharides alternate with longer regions of relatively unmodified disaccharides. The overall size of the HS chain can vary from 20 to 150 disaccharides, which adds another level of complexity to HS chains.

In the past few years, many of the enzymes that participate in HS biosynthesis have been identified on a molecular level in vertebrates and invertebrates³. In mammals, multiple isoforms of HS biosynthetic enzymes have been identified⁴. Similarly, many of the same genes have been characterized in *Drosophila*, and, rather surprisingly gene redundancy is minimal in this organism, multiple polymerases and sulphotransferases have now been identified (Table 1).

Core proteins involved in HSPG biosynthesis

A variable proportion of some proteins can become glycanated with HS chain(s) (CD-44, betaglycan) but we will restrict our discussion to core proteins that are consistently and fully glycanated with HS chains. These HSPGs are diverse and can be transmembrane (syndecans), bound by a glycosyl phosphatidylinositol (GPI) linkage to plasma membrane lipid (glypicans), or secreted into basement membranes (perlecan, agrin)⁸ (Fig. 2). Glypicans and syndecans represent the two main cell-surface HSPGs, and in mammals four syndecan and six glypican separate genes have been identified. In mice, these generally show distinct developmental and tissue-expression patterns. In *Drosophila*, only a single syndecan gene is known, but two glypican genes, *dally*⁹ and *dally-like* (G. Baeg, X. Lin and N.P., unpublished data) have been characterized (Table 1). The *unc-52* gene is the perlecan ortholog in *Caenorhabditis elegans*, and although basement membrane HSPGs have not been described in *Drosophila*, the fly genome does contain a sequence corresponding to a perlecan-like core protein.

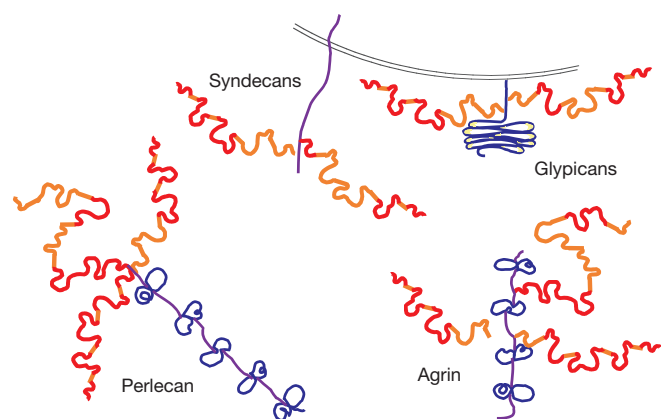


Figure 2 Depiction of HSPGs associated with cell surfaces. The syndecan and glypican family members are integral membrane proteins. The syndecan core proteins are apparently highly extended transmembrane proteins that contain a short (34–38-residue) carboxy-terminal cytoplasmic domain. The HS chains on the syndecans are linked to serine residues that are distal from the plasma membrane. The glypican core proteins are apparently disulphide-stabilized globular core proteins linked to the plasma membrane by an apparently extended C-terminal region that culminates in a GPI linkage. The HS chains on the glypicans are linked to serine residues adjacent to the plasma membrane in the apparently extended region. Basement membranes in mammalian tissues contain perlecan and agrin. Perlecan is in nearly all basement membranes as well as in cartilage, whereas agrin is a component of basement membranes in muscle, myoneural junctions, the central nervous system, lung and kidney. Both are large multidomain proteins bearing HS chains near their amino termini. Although both proteins exist as multiple isoforms that arise by alternative splicing and contain globular domains that are homologous to laminin G and epidermal growth factor domains, on the basis of their genomic organization, they are quite distinct proteins.

Although syndecans and glypicans are associated with the plasma membrane, it is important to note that under certain conditions the syndecan core proteins can be proteolytically cleaved near the cell surface, releasing the intact ectodomain replete with all the HS chains of the parental molecules (see ref. 2). Glypicans may also be shed, as they are reported to be soluble components of conditioned cell-culture media and follicular fluid.

Although there is no evidence for regulated shedding or shedding *in vivo* (see ref. 10), this shedding has been suggested to be mediated by a phosphoinositide-specific phospholipase C. In mammalian cells, syndecan ectodomain shedding is mediated by cell-surface-associated metalloproteinase(s), is regulated by intracellular signalling pathways, and results in soluble effectors¹¹. Soluble syndecan ectodomains are found in wound fluids¹² where they can inhibit growth factor¹³ and enhance proteinase activities¹⁴. Interestingly, enhanced shedding in the skin wounds of transgenic mice over-expressing syndecan-1 results in inhibition of cell proliferation and altered proteolytic activity in the wound, causing defective repair (V. Kainulainen, O. Reizes, J. Madri and M.B., unpublished data). Thus, the soluble ectodomains can act as dominant-negative inhibitors of the cell-surface HSPGs. In principle, the cleavage of the extracellular domain of HSPGs could be associated with other functions; for example, the solubilized ectodomain could regulate the movement of growth factors within extracellular spaces.

Developmental roles of HS enzymes

The critical roles of HSPGs in developmental processes and specific signalling pathways have been illustrated by the identification of mutations in enzymes involved in HS chain biosynthesis in *Drosophila* and mice.

In *Drosophila*, mutations in the homologue of UDP-D-glucose dehydrogenase, which produces UDPGlcA (*sugarless* (*sgl*)^{15–17}, and in the NDST homologue (*sulfateless* (*sfl*)¹⁸ have been identified. In either *sgl* or *sfl* mutant embryos, Wingless (Wg), a member of the Wnt family, fibroblast growth factor (FGF) and Hedgehog (Hh) signalling pathways are impaired. Interestingly, the phenotype associated with these mutations is very severe and corresponds to a complete loss of activity of these pathways, indicating that the HSPGs are absolutely required for signalling of these growth factors and morphogens.

The *Drosophila* gene *tout velu* (*ttv*) has been identified as an enzyme involved in the production of HS chains. Surprisingly, the *ttv* mutant only affects a specific signalling pathway, suggesting an unexpected level of specificity of the HSPG involved. Mutations in the *ttv* gene were first isolated based on the inability of Hh molecules to move through a field of cells¹⁹. Further studies suggested that the membrane-targeted cholesterol-modified Hh molecule requires HSPGs either to be trapped by receiving cells or to move from cells to cells²⁰. Molecular characterization of *ttv* revealed that it is

Table 1 Core proteins and enzymes

	<i>Drosophila</i>	Mice
Core proteins		
Glypican	Dally, Dally-like*	Glypican-1, -2, -3, -4, -5, -6*
Syndecan	Dsyndecan	Syndecan-1, -2, -3, -4
Enzymes		
UDP-D-glucose dehydrogenase	Sugarless	
N-deacetylase/N-sulphotransferase	Sulfateless	NDST-1, -2, -3, 4
GlcNAc/GlcA polymerase	Tout velu, DExt2?	Ext 1, Ext2, EXTL1?†
α-GlcNAc transferase?		EXTL2
Uronosyl C5 epimerase		Ep-1
2-O-Sulphotransferase	Pipe, HS2-OST	2-OST-1
3-O-Sulphotransferase		3-OST-1, -2, -3
6-O-Sulphotransferase		6-OST-1, -2, -3, -4, -5, -6, -7

* Although *Drosophila* Dally and Syndecan are known to contain HS chains, it is important to note that in most instances the enzymatic activities of the proteins listed have not been shown and are inferred from their amino-acid sequence homologies.

† Proteins of the EXT/EXTL families encode either HS-copolymerases (Fig. 1) or α-GlcNAc transferases (the enzyme that begins the HS chain after linker biosynthesis²⁰). In particular, there is no direct evidence that *Sgl* and *Sfl* have the enzymatic activities listed, and the biochemical activities of DExt2 or mouse EXTL1 have not yet been reported.

homologous to a member of the *Ext* gene family, which has been implicated in the human multiple exostoses (Ext) syndrome²¹. This dominantly inherited disease is characterized by short stature, inequalities in limb length, bone deformities and the presence of bone outgrowths, called exostoses, at the ends of long bones. The function of Ext proteins as HS polymerases has recently been elucidated. First, the inability of herpes simplex virus, which binds to cell-surface HS, to infect the mouse cell line sog9 correlates with a defect in HS biosynthesis²². This defect, however, can be rescued by introducing an *Ext1* complementary DNA into the mutant cell line. Second, an *Ext2* homologue from bovine serum has been isolated²³ and shown to have HS polymerase activity.

Because *ttv* encodes a putative HS polymerase, it is surprising that in the absence of *ttv* activity only Hh signalling is affected, and not Wg and FGF signalling as observed in *sgl* and *sfl* mutants. This, together with the observation that the overall HS concentration is reduced in *ttv* mutants^{20,24}, suggests that other *Drosophila Ext* genes exist. Consistent with this finding, at least one other *Ext* gene is present in the fly genome²⁰. A possible explanation of the specificity of Ttv for Hh action is that Hh signalling may be more sensitive to a reduction in HSPG concentration than Wg and FGF signalling. Alternatively, Hh-specific HSPG(s) may exist and Ttv may be responsible for its production.

In addition to Ttv/Ext, *Drosophila* mutants in the *pipe* gene, which encodes a putative 2-O-sulphotransferase²⁵, show defects in the establishment of dorsal–ventral (D–V) polarity in the embryo. *pipe* is expressed in the ventral part of the egg chamber and has been proposed to activate the serine protease cascade that leads to production of the active Toll ligand Spätzle. The effect of *pipe* on Toll signalling is probably indirect. The role of the proteoglycan involved might be either to activate or to assemble a protease complex which processes Spätzle or concentrates Spätzle to the ventral side of the egg chamber. Finally, although Pipe encodes a putative 2-OST, it is not yet known whether it encodes a sulphotransferase for a HSPG or another class of proteoglycan.

Similar examples of specific effects associated with enzymes involved in HS chain biosynthesis have also been found in mice. Compelling evidence for tissue-type selectivity of HSPG function comes from studies of embryos that are homozygous for a mutation in the gene encoding HS 2-OST²⁶. These mice show multiple developmental abnormalities including lack of kidney induction. The abnormalities probably arise from compromised activities of the specific HS-dependent growth factors responsible for this organogenesis. In addition, targeted disruption of a single gene encoding one of four known NDST genes prevents mast-cell biosynthesis of heparin chains but does not affect the biosynthesis of other HS chains^{27,28}. Thus, the HS biosynthetic enzymes probably have activities that are specific to cell type. Differences in the activities or regulation of these enzyme isoforms may account for the reproducible structural distinctions of HS chains from different cell and tissue types^{29,30} and for reproducible differences in growth factor binding of HS chains from distinct developmental stages³¹.

The HSPG core proteins have specific developmental roles

In contrast to the analysis of the HS biosynthetic enzymes, much less is known about the phenotypes associated with mutations in the core proteins. In *Drosophila*, mutations in both *syndecan* and one of the *glypican* genes (*dally*) have been identified⁹; however, detailed phenotypic analysis are only available for *dally*.

Genetic analyses of *dally* have implicated this glypican in both Wg and Decapentaplegic (Dpp, a member of the transformin growth factor- β family) signalling. Loss of *dally* activity, in both the embryo and imaginal discs, generates phenotypes reminiscent of loss of Wg activity. Genetic experiments are consistent with a model in which Dally acts as a co-receptor for the Wg-transducing receptor encoded by the seven-transmembrane protein Frizzled 2 (refs 18, 32). Interestingly, *dally* expression is developmentally regulated and is

co-expressed with Frizzled 2 in the embryo, suggesting that, as previously observed for Frizzled 2 (ref. 33), the level of the HSPG co-receptor has to be tightly regulated for proper Wg signalling.

Interestingly, Dally has been found to act in Dpp signalling but only in imaginal discs³⁴. Dpp is also expressed during embryonic stages; however, there is no evidence from the studies of *sfl*, *sgl* or *dally* mutants that HSPGs are involved in the early function of Dpp in the establishment of D–V embryonic polarity. A reduction in *dpp* levels enhances the defects associated with *dally* mutations in the eye, antenna and genitalia. Furthermore, additional copies of *dpp* rescue the defects in these tissues. These genetic interactions have led to the hypothesis that Dally regulates Dpp activity³². Analysis of Dally provides a striking example of an HSPG that is developmentally regulated and that interacts with only a subset of signalling pathways.

In mice, targeted disruptions of glypican genes yield either no apparent abnormalities (glypican-2; S. Saunders and A. Lander, personal communication) or a phenotype similar to the Simpson–Golabi–Behmel syndrome³⁵, a congenital overgrowth syndrome of unknown pathogenesis in humans associated with mutations in the *glypican-3* gene³⁶. Interestingly, glypican-3 may function to regulate bone morphogenetic protein (BMP)-4 signalling in as much as the progeny of crosses between the glypican-3 knockout with haplo-insufficient mutants of BMP-4 in mice show skeletal abnormalities that are not seen in either parental strain (S. Paine-Saunders, W. C. Skarnes and S. Saunders, personal communication). Targeted disruption of the *syndecan-1* gene, the main syndecan gene expressed by epithelia and plasma cells, yields apparently normal fertile mice despite the early and intense expression of syndecan-1 during embryogenesis¹. The absence of abnormalities presumably results from compensatory expression of other syndecan family members. However, epidermal or corneal epithelial wounds in these mice undergo markedly delayed repair (H. Gibson, M. Hinkes and M.B., unpublished data; M. A. Stepp *et al.*, unpublished data). Furthermore, mammary epithelia in these mice resist the tumorigenic action of the transgenically expressed Wnt-1 oncogene³⁷ and also resist lung infection when certain bacteria are inoculated intranasally, but are normally susceptible when inoculated via the peritoneal cavity (P. W. Park, G. Pier and M.B., unpublished data). Thus, abnormalities are detected in these syndecan-1-deficient mice only by perturbing their epithelia. These data suggest that compensation by other HSPGs is sufficient to maintain normal development and reproduction, but cannot replace the specific functions of syndecan-1.

Perspectives

We have discussed the contribution of the genetic approach to the understanding of HSPG functions in developmental processes. It is now apparent that analysis of the mutant phenotypes associated with specific gene disruptions in either the various enzymes involved in HS chain biosynthesis or the core proteins provide a wealth of information about the complex roles of HSPGs *in vivo*. To gain further insights into these roles, it is now critical to analyse the mutant phenotypes associated with the remaining enzymes and core proteins. The availability of a repertoire of mutants in each of the components involved in HS chain biosynthesis will allow many other issues to be critically addressed. In particular, detailed analysis of HS chain structure in various mutants will define the physiological roles of HS chains expressing different binding specificities. Furthermore, issues of functional redundancy between HSPG core proteins should be directly testable by analysing the phenotypes of double mutants, or by expressing one specific HSPG core protein in the mutant background of another.

The availability of HSPG mutants will allow one to characterize the function of HSPGs in signalling. In most models the HSPG is viewed as a co-receptor for the signal-transducing receptor; however, in some instances the extracellular domain can be cleaved,

released from the cell surface and potentially used as a carrier of extracellular signals. Using genetic mosaics, the cellular autonomy of HSPG function could be readily determined. In the case of Hh signalling, the HSPG appears to be critical to the cell-to-cell movement of Hh molecules. Although the precise roles of the HSPGs are not yet understood, it is clear that these complex molecules are fundamental to the distribution and processing of extracellular signals that pattern fields of cells. Finally, additional roles of HSPGs in developmental, physiological and behavioural processes are likely to be uncovered by the analysis of mutants. □

Received 5 November 1999; accepted 10 February 2000.

- Bernfield, M. *et al.* Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu. Rev. Cell Biol.* **8**, 365–393 (1992).
- Bernfield, M. *et al.* Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* **68**, 729–777 (1999).
- Habuchi, H., Habuchi, O. & Kimata, K. Biosynthesis of heparan sulfate and heparin. How are the multifunctional glycosaminoglycans built up? *Trends Glycosci. Glycotechnol.* **10**, 65–80 (1998).
- Rosenberg, R. D., Shworak, N. W., Liu, J., Schwartz, J. J. & Zhang, L. Heparan sulfate proteoglycans of the cardiovascular system. Specific structures emerge but how is synthesis regulated? *J. Clin. Invest.* **99**, 2062–2072 (1997).
- Lindahl, U., Kusche-Gullberg, M. & Kjellén, L. Regulated diversity of heparan sulfate. *J. Biol. Chem.* **273**, 24979–24982 (1998).
- Lyon, M. & Gallagher, J. T. Bio-specific sequences and domains in heparan sulfate and the regulation of cell growth and adhesion. *Matrix Biol.* **17**, 485–493 (1998).
- Lindahl, U. *Heparin* (CRC, Boca Raton, Florida, 1989).
- Iozzo, R. V. Matrix proteoglycans: from molecular design to cellular function. *Annu. Rev. Biochem.* **67**, 609–652 (1998).
- Nakato, H., Futch, T. A. & Selleck, S. B. The division abnormally delayed (*dally*) gene: a putative integral membrane proteoglycan required for cell division patterning during postembryonic development of the nervous system in *Drosophila*. *Development* **121**, 3687–3702 (1995).
- Veuglers, M. & David, G. The glypicans: a family of GPI-anchored heparan sulfate proteoglycans with a potential role in the control of cell division. *Trends Glycosci. Glycotechnol.* **10**, 145–152 (1998).
- Fitzgerald, M. L. *et al.* Shedding of syndecan-1 and -4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3 sensitive metalloproteinase. *J. Cell. Biol.* **148**, 811–824 (2000).
- Subramanian, S. V., Fitzgerald, M. L. & Bernfield, M. Regulated shedding of syndecan-1 and -4 ectodomains by thrombin and growth factor activation. *J. Biol. Chem.* **272**, 14713–14720 (1997).
- Kato, M. Physiological degradation converts the soluble syndecan-1 ectodomain from an inhibitor to a potent activator of FGF-2. *Nature Med.* **4**, 691–697 (1998).
- Kainulainen, V., Wang, H., Schick, C. & Bernfield, M. Syndecans, heparan sulfate proteoglycans, maintain the proteolytic balance of acute wound fluids. *J. Biol. Chem.* **273**, 11563–11569 (1998).
- Binari, R. C. *et al.* Genetic evidence that heparin-like glycosaminoglycans are involved in *wingless* signaling. *Development* **124**, 2623–2632 (1997).
- Haecker, U., Lin, X. & Perrimon, N. The *Drosophila* sugarless gene modulates *Wingless* signaling and encodes an enzyme involved in polysaccharide. *Development* **124**, 3565–3573 (1997).
- Haerry, T. E., Heslip, T. R., Marsh, J. L. & O'Connor, M. B. Defects in glucuronate biosynthesis disrupt *Wingless* signaling in *Drosophila*. *Development* **124**, 3055–3064. (1997).
- Lin, X. & Perrimon, N. *Dally* cooperates with *Drosophila* Frizzled 2 to transduce *Wingless* signalling. *Nature* **400**, 281–284 (1999).
- Bellaiche, Y., The, I. & Perrimon, N. *Tout-velu* is a *Drosophila* homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion. *Nature* **394**, 85–88 (1998).
- The, I., Bellaiche, Y. & Perrimon, N. Evidence that heparan sulfate proteoglycans are involved in the movement of Hedgehog molecules through fields of cells. *Mol. Cell* **4**, 633–639 (1999).
- Stickens, D. *et al.* The EXT2 multiple exostoses gene defines a family of putative tumour suppressor genes. *Nature Genet.* **14**, 25–32 (1996).
- McCormick, C. *et al.* The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nature Genet.* **19**, 158–161 (1998).
- Lind, T., Tufaro, F., McCormick, C., Lindahl, U. & Lidholt, K. The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J. Biol. Chem.* **273**, 26265–26268 (1998).
- Toyoda, H., Kinoshita-Toyoda, A. & Selleck, S. B. Structural analysis of glycosaminoglycans in *Drosophila* and *C.elegans* and demonstration that *tout velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J. Biol. Chem.* **275**, 2269–2275 (2000).
- Sen, J., Goltz, J. S., Stevens, L. & Stein, D. Spatially restricted expression of *pipe* in the *Drosophila* egg chamber defines embryonic dorsal-ventral polarity. *Cell* **95**, 471–481 (1998).
- Bullock, S. L., Fletcher, J. M., Bedington, R. S. P. & Wilson, V. A. Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase. *Genes Dev.* **12**, 1894–1906 (1998).
- Forsberg, E. *et al.* Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme. *Nature* **400**, 773–776 (1999).
- Humphries, D. E. *et al.* Heparin is essential for the storage of specific granule proteases in mast cells. *Nature* **400**, 769–772 (1999).
- Kato, M., Wang, H., Bernfield, M., Gallagher, J. T. & Turnbull, J. E. Cell surface syndecan-1 on distinct cell types differs in fine structure and ligand binding of its heparan sulfate chains. *J. Biol. Chem.* **269**, 18881–18890 (1994).
- Sanderson, R. D., Turnbull, J. E., Gallagher, J. T. & Lander, A. D. Fine structure of heparan sulfate regulates syndecan-1 function and cell behavior. *J. Biol. Chem.* **269**, 13100–13106 (1994).
- Nurcombe, V., Ford, M. D., Wildschut, J. A. & Bartlett, P. F. Developmental regulation of neural response to FGF-1 and FGF-2 by heparan sulfate proteoglycan. *Science* **260**, 103–106 (1993).
- Tsuda, M. *et al.* The cell-surface proteoglycan *Dally* regulates *Wingless* signalling in *Drosophila*. *Nature* **400**, 276–280 (1999).
- Cadigan, K. M., Fish, M. P., Rulifson, E. J. & Nusse, R. *Wingless* repression of *Drosophila* frizzled 2 expression shapes the *Wingless* morphogen gradient in the wing. *Cell* **93**, 767–777 (1998).
- Jackson, S. M. *et al.* *Dally*, a *Drosophila* glypican, controls cellular responses to the TGF- β -related morphogen, Dpp. *Development* **124**, 4113–4120 (1997).
- Kato, M., Wang, H., Bernfield, M., Gallagher, J. T. & Turnbull, J. E. A glypican encoded by a gene that is mutated in the Simpson-Golabi-Behmel overgrowth syndrome, induces apoptosis in a cell line-specific manner. *J. Cell Biol.* **141**, 1407–1414 (1998).
- Pilia, G. *et al.* Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature Genet.* **12**, 241–247 (1996).
- Alexander, C. M. *et al.* Syndecan-1 is required for Wnt-1-induced mammary tumorigenesis in mice. *Nature Genet.* (in the press).
- Kitagawa, H., Shimakawa, H. & Sugahara, K. The tumor suppressor EXT-like gene EXT1-2 encodes an α 1,4-N-acetylhexosaminyltransferase that transfers N-acetylgalactosamine and N-acetylglucosamine to the common glycosaminoglycan-protein linkage region. The key enzyme for the chain initiation of heparan sulfate. *J. Biol. Chem.* **274**, 13933–13937 (1999).
- Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).

Acknowledgements

This work was supported by the NIH (MB). N.P. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to N.P. or M.B.

One-dimensional nature of the magnetic fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$

H. A. Mook*, Pengcheng Dai*, F. Doğan† & R. D. Hunt*

* Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

† Department of Materials Sciences and Engineering, University of Washington, Seattle, Washington 98195, USA

There is increasing evidence that inhomogeneous distributions of charge and spin—so-called ‘striped phases’—play an important role in determining the properties of the high-temperature superconductors. For example, recent neutron-scattering measurements on the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ family of materials show both spin and charge fluctuations that are consistent with the striped-phase picture. But the fluctuations associated with a striped phase are expected to be one-dimensional, whereas the magnetic fluctuations observed to date appear to display two-dimensional symmetry. We show here that this apparent two-dimensionality results from measurements on twinned crystals, and that similar measurements on substantially detwinned crystals of $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ reveal the one-dimensional character of the magnetic fluctuations, thus greatly strengthening the striped-phase interpretation. Moreover, our results also suggest that superconductivity originates in charge stripes that extend along the *b* crystal axis, where the superfluid density is found to be substantially larger than for the *a* direction.

Striped phases are inhomogeneous distributions of charge and spin that have been suggested to account for many of the unusual properties of the high- T_c copper oxide superconductors^{1–7} (T_c is the superconducting transition temperature). In the simplest picture, the charge and spin can be thought of as being confined to separate linear regions in the crystal and thus resembling stripes. We expect static striped phases for the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ materials not to coexist with superconductivity, except perhaps for materials with low oxygen contents and low values of T_c . However, neutron-scattering measurements have shown results for fluctuations of both spin^{8,9} and charge¹⁰ that support the existence of a dynamic striped phase in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ materials that have high T_c values. Nevertheless, a difficulty for the dynamic striped-phase picture is that, for both $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (refs 8, 9) and $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (refs 11, 12) copper oxide superconductors, the magnetic fluctuations stemming from the spins have displayed a four-fold pattern at incommensurate points around the magnetic (1/2, 1/2) reciprocal lattice position, as shown in Fig. 1A. The interpretation of these measurements has been unclear, with both striped phases^{1–7} and nested Fermi surfaces^{13–15} being possible explanations. Because the striped phase is expected to propagate along a single direction, only a single set of satellites around the antiferromagnetic position should be observed. Tranquada^{7,16} has suggested that the four-fold symmetry might arise from stripes that alternate in direction as the planes are stacked along the *c* axis.

The problem in interpreting the four-fold pattern stems from the fact that the crystals used in the neutron experiments are twinned, so that along a given *a* or *b* direction, domains with lattice spacing *a* or *b* exist in equal proportion. This makes it impossible to distinguish whether satellites originate from the *a** or *b** direction in the reciprocal lattice. What is needed to resolve the problem of interpreting the magnetic measurements is either an untwinned crystal, or a crystal with a substantial difference in the domain population along the different directions in the *a*–*b* plane.

Sample preparation

We have attempted to detwin $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ crystals by applying a high pressure to one of the *a*(*b*) directions of a crystal during oxygenation, so that as the crystal undergoes the tetragonal to orthorhombic transition, the squeezed direction prefers to be the lower-lattice-constant *a* direction. Crystals are commonly detwinned in this manner, but it is challenging to accomplish this with a crystal large enough to observe the magnetic incommensu-

rate scattering with neutrons. Typically, either the crystal breaks or only a small difference in domain population is obtained. The 4.39-gram crystal used in the neutron measurements was clamped vertically, with the *c* axis perpendicular to the load direction, between two aluminium oxide ceramic rods that were attached to the water-cooled plates of a hydraulic press. A thin layer of refractory wool was placed between the sample and rods to reduce any possible non-uniform stress distribution under applied pressure, and to enhance oxygen diffusion along the clamped surfaces. Oxygenation of the sample took place in a tube furnace at 400°C while a constant pressure of 13.9 MPa was applied for 120 hours in a flowing-oxygen gas environment. The sample was then reduced to an oxygen content of about 6.6, in accordance with the techniques developed during an extensive experimental study¹⁷ on the nonstoichiometry in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ (an oxygen content of 6.6 indicates that *x* in this formula is 0.4). An equation was developed in this study relating the oxygen content and T_c to the partial oxygen pressure and temperature. The detwinning process resulted in a sample in which the domain population has roughly a 2:1 ratio.

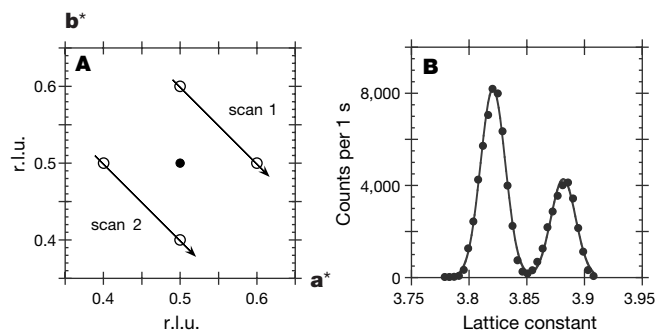


Figure 1 Diagram showing the scans used in the experiment, and a plot of the twin domain population. **A**, Open circles show the positions in reciprocal lattice units (r.l.u.) of the magnetic incommensurate satellites that surround the antiferromagnetic position shown by the filled circle. The arrows show the two scans used in the experiment. The centre of scan 1 is at (0.55, 0.55, 2) in r.l.u., while the centre of scan 2 is (0.45, 0.45, 2) r.l.u. **B**, Scan for the *a** direction obtained from the (2, 0, 0) reflection given in units of the lattice constant for the partially detwinned sample. The lattice constants are those expected for an oxygen concentration near 6.6. The peak heights give the domain population of the partially detwinned sample.

The Bragg scattering from one of the **a**(**b**) directions from the crystal is shown in Fig. 1B. The pattern was taken with the HB-3 spectrometer at the High-Flux Isotope Reactor (HFIR) in Oak Ridge, Tennessee, USA, using 14.7-meV neutrons obtained from the (002) reflection of a Be monochromator. The angular collimation used was 10 minutes before and after the sample. The peak corresponding to lattice constant *a* is considerably bigger than that for the lattice constant *b*, so we denote this face as the **a** face of the crystal. Least-squares fitting a gaussian distribution to the scattering indicates a domain ratio (*a*/*b*) of 1.92 ± 0.05 . Measurements on the **b** face confirmed this result with an identical domain ratio for (*b*/*a*) within the error of the measurement.

The pattern of the magnetic fluctuations

We then modified the spectrometer configuration so that we could observe the incommensurate magnetic fluctuations for both the **a*** and **b*** directions in the reciprocal lattice of the crystal. Pyrolytic graphite monochromator and analyser crystals were installed on the HB-3 triple-axis spectrometer, and the spectrometer was set for an energy transfer of 24 meV using a final energy of 30.5 meV. Collimations were 48–40–60–120 minutes from in front of the monochromator until after the analyser. Measurements were first made on the 25.6-gram twinned crystal used to originally determine the location of the incommensurate scattering peaks⁹. The crystal was oriented so that the scans through the satellites could be made as shown in Fig. 1A while keeping **c*** set at 2 r.l.u. (reciprocal lattice units) so as to maximize the bilayer structure factor^{8,9}. These scans have the advantage that the neutron detector remains at a nearly constant angle, thus avoiding background contamination from the direct neutron beam of the triple-axis spectrometer. This permits a measurement at the small total momentum transfer used, which maximizes the magnetic form factor while minimizing the effects of phonon scattering. For YBa₂Cu₃O_{6.6} the satellites are found in the twinned crystal at approximately 0.1 r.l.u. along **a*** and **b*** from the magnetic superlattice peak. The scan shown in Fig. 2A is denoted scan 1 in Fig. 1A. The centre of this scan is at 0.55 r.l.u. while the

peaks are expected at 0.5 and 0.6 r.l.u. along **a***. The scan in Fig. 2B is denoted scan 2 in Fig. 1A. The scan centre is at 0.45 r.l.u. in this case, with the peaks expected at 0.4 and 0.5 r.l.u. along **a***. Least-squares fitting of gaussian distributions yields a ratio of the intensities of the satellites along **a*** to those along **b*** of 1.05 ± 0.05 , which is consistent with the equal intensities for the two directions expected for the completely twinned crystal.

Measurements for the partially detwinned sample are shown in Fig. 2C, D. Least-squares fitting in the same manner for these data results in a ratio of the satellite intensities along **a*** to those along **b*** of 1.95 ± 0.21 , which is the same within the error bars as the population of the domains for the two directions. Thus, it follows within an error of 11% that 100% of the satellites stem only from the **a*** domain, showing that the spin fluctuations have a one-dimensional (1D) modulation along **a**. A 1D static magnetic modulation along **b** has recently been found by Wakimoto *et al.*¹⁸ for the insulating material La_{1.95}Sr_{0.05}CuO₄. This suggests the possibility of a static striped phase propagating along **b**, although in this case no evidence of charge scattering was identified.

Phonon measurements

Now that we have discovered that the incommensurate magnetic scattering has a 1D modulation vector for YBa₂Cu₃O_{6.6}, it is worthwhile revisiting the measurements of the charge scattering as imaged by the phonons¹⁰. Because the striped-phase modulation vector occurs along **a**, the phonons from the **b** direction should display normal behaviour. In a twinned crystal, the charge scattering measured in the **a***(**b***) direction should have one component that has normal behaviour and one component that is strongly affected by the striped phase. In fact, the measured line shapes shown in figure 2 of ref. 10 give an indication that the wide peak where the charge anomaly is found may consist of two parts. One of the parts would be expected to be sharp and occur at the unperturbed phonon position, and the other would be strongly damped by the charge fluctuations. Calculations of the phonon line shapes in striped-phase materials show a rather complicated line shape resulting from the interaction of the charge fluctuations with the phonons (N. Wakabayashi *et al.*, manuscript in preparation). Nevertheless, the knowledge that the striped-phase modulation vector occurs mostly along **a** justifies a higher-resolution study of the phonon anomaly.

Figure 3 shows a high-resolution measurement at the 4.25 r.l.u. position for the optical phonon studied in ref. 10 for YBa₂Cu₃O_{6.6}

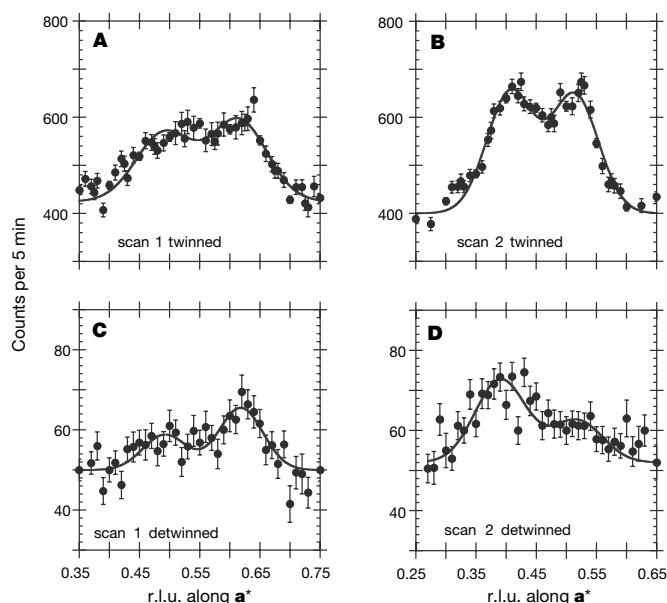


Figure 2 Scans of the magnetic incommensurate scattering for the twinned and partially detwinned sample taken at 10 K. The scan directions are shown in Fig. 1A. The counting errors shown were obtained by averaging a number of runs together. The larger magnetic scattering for scan 2 is accounted for by the Cu magnetic form factor. The scan is made in the diagonal direction shown, with the scale on the graphs giving the component along **a***.

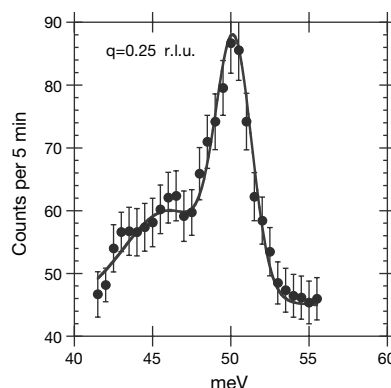


Figure 3 Optical phonon measurements on YBa₂Cu₃O_{6.6}. The figure shows a high-resolution measurement at 10 K of the optical phonon discussed in ref. 10. The measurement was made to image the charge fluctuations in YBa₂Cu₃O_{6.6}. The higher resolution makes it clearer that the broad line shape resulting from the charge fluctuations can be interpreted to consist of the superposition of two lines; a normal narrow line from the **b** twin domain, and a broad line from the **a** twin that contains the striped-phase 1D modulation vector.

using the twinned crystal. This is in the region of momentum space where we expect the phonons to be strongly affected by the striped phase. The fit shown is for two gaussian distributions. The position of the narrow peak occurs at 50.1 meV, consistent with the value found in lightly doped materials where any striped phase would have a much smaller wavevector, while the width of 2.9 meV is the same as the experimental resolution. The other distribution is found to have a wide width of 5.7 meV, as might be expected from the interaction with a stripe-phase dynamic charge density wave. The areas under the gaussian distributions are equal within experimental errors. This is consistent with our result for the magnetic scattering, and indicates a method of determining whether the charge fluctuations are 1D or 2D.

Stripes versus Fermi liquid

The Fermi liquid concept, with its concomitant quasiparticles and Fermi surface, has been successful in solving many of the important problems in condensed-matter physics, including conventional superconductivity. This approach for the layered copper oxide superconductors results in a 2D Fermi surface. We are unaware of any Fermi surface calculation that can account for our measured results unless 1D effects are somehow incorporated into the calculation. The stripe model takes a completely different approach, wherein the electronic matter self-organizes into a topologically 1D-quantum texture. Such a concept naturally incorporates the 1D modulation observed in the present experiment. This provides compelling evidence that the observed spin and charge fluctuations stem from a dynamic striped phase. We do not have sufficient information to understand in detail the characteristics of such a phase. Nevertheless, such a phase is a novel state of matter decidedly different from those stemming from the conventional Fermi-liquid ideas.

However, the striped-phase model is not able to completely characterize the observed magnetic fluctuations. Much of the weight of the total magnetic scattering is found in the commensurate resonance excitation^{19–21}. So far, it is unclear how the striped-phase concept can account for the resonance. It is also uncertain if the striped phase has any direct connection to the superconductivity. It may be, though, that the 1D nature of the striped phase simplifies the difficulties encountered in understanding the nature of the superconducting pairing mechanism²². The charge stripes extend along **b**, which is in the same direction as the Cu–O chains found in the YBa₂Cu₃O_{7–x} structure. If superconductivity occurs in the charge stripes, anisotropic superconducting behaviour should be found in the **a**–**b** plane. Far-infrared spectroscopy measurements by Basov *et al.*^{23,24} show that the London penetration depth is considerably smaller, or the superfluid density larger, along the **b** direction for the fully doped material. This was attributed to a substantial portion of the superconducting condensate residing on the chains. However, the superfluid density was also found to be about 1.7 times larger for the underdoped material YBa₂Cu₃O_{6.6},

where the missing atoms in the chains would be expected to inhibit the chain contribution to the superconductivity. The present measurements show that the stripes extend in the correct direction to account for the observed anisotropy in the superfluid density. □

Received 4 January; accepted 14 March 2000.

1. Zaanen, J. & Gunnarson, O. Charged magnetic domain lines and the magnetism of high-*T_c* oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
2. Poilblanc, D. & Rice, T. M. Charged solitons in the Hartree-Fock approximation to the large-*U* Hubbard model. *Phys. Rev. B* **39**, 9749–9752 (1989).
3. Schulz, H. J. Domain walls in a doped antiferromagnet. *J. Phys. (Paris)* **50**, 2833–2849 (1989).
4. Kato, M., Machida, K., Nakanishi, H. & Fujita, M. Soliton lattice modulation of incommensurate spin density wave in two-dimensional Hubbard model—a mean field study. *J. Phys. Soc. Jpn* **59**, 1047–1058 (1990).
5. Kivelson, S. A. & Emery, V. J. in *Strongly Correlated Electronic Materials: The Los Alamos Symposium 1993* (eds Bedell, K. S. *et al.*) 619–656 (Addison-Wesley, Maine, 1994).
6. Emery, V. J. & Kivelson, S. A. Collective charge transport in high-temperature superconductors. *Physica C* **235–240**, 189–192 (1994).
7. Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
8. Dai, P., Mook, H. A. & Doğan, F. Incommensurate magnetic fluctuations in YBCO_{6.6}. *Phys. Rev. Lett.* **80**, 1738–1741 (1998).
9. Mook, H. A. *et al.* Spin fluctuations in YBa₂Cu₃O_{6.6}. *Nature* **395**, 580–582 (1998).
10. Mook, H. A. & Doğan, F. Charge fluctuations in YBa₂Cu₃O_{7–x} high-temperature superconductors. *Nature* **401**, 145–147 (1999).
11. Cheong, S.-W. *et al.* Incommensurate magnetic fluctuations in La_{1–x}Sr_xCuO₄. *Phys. Rev. Lett.* **67**, 1791–1794 (1991).
12. Mason, T. E., Aeppli, G. & Mook, H. A. Magnetic dynamics of superconducting La_{1–x}Sr_xCuO₄. *Phys. Rev. Lett.* **68**, 1414–1417 (1992).
13. Brinckmann, J. & Lee, P. A. Slave boson approach to neutron scattering in YBa₂Cu₃O_{6+y} superconductors. *Phys. Rev. Lett.* **82**, 2915–2918 (1999).
14. Kao, Y.-J., Si, Q. & Levin, K. Frequency evolution of neutron peaks below *T_c*: commensurate and incommensurate structure in LaSrCuO and YBaCuO. Preprint cond-mat/9908302 at <http://xxx.lanl.gov> (1999; cited 22 Aug. 1999).
15. Norman, M. R. The relation of neutron incommensurability to electronic structure in high-temperature superconductors. Preprint cond-mat/9912203 at <http://xxx.lanl.gov> (1999; cited 13 Dec. 1999).
16. Tranquada, J. M. Modulated spin and charge densities in cuprate superconductors. *Physica B* **241–243** (1998).
17. Lindemer, T. B. *et al.* Experimental and thermodynamic study of nonstoichiometry in YBa₂Cu₃O_{7–x}. *J. Am. Ceram. Soc.* **72**, 1775–1788 (1989).
18. Wakimoto, S. *et al.* Direct observation of a one-dimensional static spin modulation in insulating La_{1.95}Sr_{0.05}CuO₄. *Phys. Rev. B* **61**, 3699–3706 (2000).
19. Rossat-Mignod, J. *et al.* Neutron scattering study of the YBa₂Cu₃O_{6+x} system. *Physica C* **185**, 86–92 (1991).
20. Mook, H. A., Yethiraj, Y., Aeppli, G., Mason, T. E. & Armstrong, T. Polarized neutron determination of the magnetic excitations in YBa₂Cu₃O₇. *Phys. Rev. Lett.* **70**, 3490–3493 (1993).
21. Fong, H. *et al.* Phonon and magnetic neutron scattering at 41 meV in YBa₂Cu₃O₇. *Phys. Rev. Lett.* **75**, 316–319 (1995).
22. Emery, V. J., Kivelson, S. A. & Zahar, O. Spin-gap proximity effect mechanism of high-temperature superconductivity. *Phys. Rev. B* **56**, 6120–6147 (1997).
23. Basov, D. N. *et al.* In-plane anisotropy of the penetration depth in YBa₂Cu₃O_{7–x} and YBa₂Cu₄O₈ superconductors. *Phys. Rev. Lett.* **74**, 598–601 (1995).
24. Hardy, W. N. *et al.* in *Proc. 10th Anniversary HTS Workshop on Physics, Materials and Applications 1996* (eds Batlogg B. *et al.*) 223–227 (World Scientific, Singapore, 1996).

Acknowledgements

The work at Oak Ridge National Laboratory was supported by the US Department of Energy.

Correspondence and requests for materials should be addressed to H.A.M. (e-mail: ham@ornl.gov).

Magnetic field surrounding the starburst nucleus of the galaxy M82 from polarized dust emission

J. S. Greaves, W. S. Holland, T. Jenness & T. G. Hawarden

Joint Astronomy Centre, 660 N. A'ohōkū Place, Hilo, Hawaii 96720, USA

Magnetic fields may play an important role in the star-formation process, especially in the central regions of 'starburst' galaxies where star formation is vigorous. But the field directions are very difficult to determine in the dense molecular gas out of which the stars form, so it has hitherto been impossible to test this hypothesis. Dust grains in interstellar clouds tend to be magnetically aligned, and it is possible to determine the alignment direction based on the polarization of optical light due to preferential extinction along the long axes of the aligned grains¹. This technique works, however, only for diffuse gas, not for the dense molecular gas. Here we report observations of polarized thermal emission from the aligned dust grains in the central region of M82, which directly traces² the magnetic field structure (as projected onto the plane of the sky). Organized field lines are seen around the brightest star-forming regions, while in the dusty halo the field lines form a giant magnetic bubble possibly blown out by the galaxy's 'superwind'.

Submillimetre polarimetry is a relatively new technique for mapping astronomical magnetic fields. Elongated dust grains in interstellar environments can be partially aligned by internal paramagnetic relaxation², spinning with their long axes perpendicular to the field. Thermal emission is then preferentially enhanced along the

predominant grain orientation at levels of a few per cent. A number of Galactic star-forming clouds have been imaged with submillimetre polarimetry^{3,4}, but such studies have previously been too technically challenging to perform for other galaxies.

M82 is advantageous in being a nearby starburst (approximate distance of 3.25 Mpc), and the brightest region has a flux density at 850 μm of 1.4 Jy in the 15-arcsec beam of the 15-m James Clerk Maxwell Telescope (JCMT), comparable in surface brightness to moderate star-forming environments in the solar neighbourhood⁵. The main features of the submillimetre dust emission of M82 are two peaks symmetric about the galaxy's dynamical centre⁶, variously interpreted as the ends of a highly inclined torus, or of a bar⁷. The starburst region contains at least 100 stellar superclusters inside the dust zone⁸, and high-speed energetic winds are observed emerging perpendicular to the major axis, expelling grains into a halo extending more than 1 kpc above the galactic plane⁹. Radio polarimetry data¹⁰ have shown a poloidal field within the wind, at heights greater than 400 pc, and infrared polarimetry¹¹ reveals that this direction originates at the nucleus, but the magnetic structure in the dusty torus has been hitherto undetectable.

To observe the magnetic fields in the starburst region, we used the SCUBA camera¹² and its polarimeter^{13,14} on the JCMT, observing over three nights in February 1998 and April 1999. The results are shown in Fig. 1, with vectors depicting the magnetic field directions as projected onto the plane of the sky, superimposed on an image of the dust emission.

Organized magnetic field structure has been detected, to our knowledge for the first time, around a starburst galaxy nucleus. In M82, two main components are visible, in the circumnuclear torus and extended halo regions respectively. Towards the bright peaks, the field is very ordered (black vectors) but the directions differ for the east and west sides of the torus. The field orientations are respectively 40° and 130° east of north, with dispersions of only ± 5 –10° for the pairs of vectors nearest the flux peaks. (Because the telescope beam size is 15 arcsec, the points 9–12 arcsec apart in each pair sample largely independent lines of sight.) These closely aligned

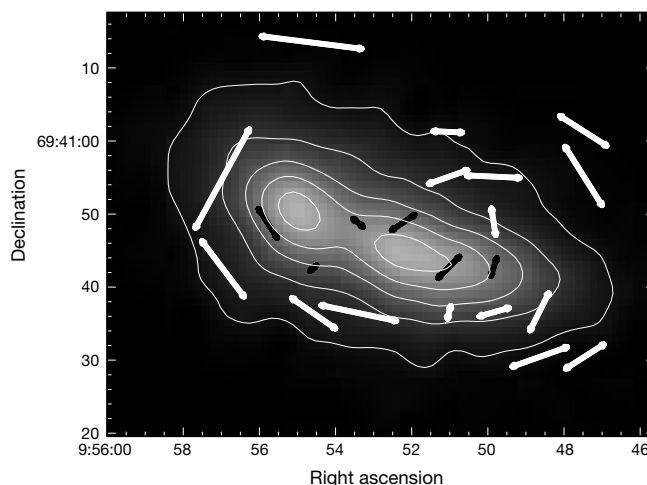


Figure 1 Polarimetry results for M82. The vectors show the inferred magnetic field directions (rotated 90° from the observed electric field plane of the polarized 850- μm emission), and the lengths are proportional to polarization percentage. The greyscale is an image of 450- μm dust emission (with negligible contributions from other emission mechanisms⁹), obtained simultaneously with the 850- μm polarization data, and smoothed to a spatial resolution of 12 arcsec (850- μm emission appears similar but with the peaks less well resolved). The contours are at intervals of 10, 30, 50, 70 and 90% of the 450- μm peak flux of approximately 5 Jy per beam. Vectors above the 50% flux level are shown in black, and those below it in white, for clarity. The range of polarization levels is 0.7–9.1% and the vectors are plotted at independently sampled points 6 arcsec apart.

All vectors shown have a significance level greater than 3σ , with a median value of 6σ defining the position angles to $\pm 5^\circ$. The datasets from February 1998 and April 1999 were compared and vectors with significant discrepancies were rejected (even though the earlier data were twice as noisy), to ensure that the final results are real. The data are reduced by subtracting instrumental polarization and fitting the sinusoidally modulated signal¹³, the sky background was removed by chopping the telescope secondary mirror over 2 arcmin, while waveplate steps of 22.5° and switching between left and right beams were performed every 0.5–2 minutes. The total integration time was 3.7 hours, enabling detections of polarization within the 10% flux contour at 850 μm ; in fainter regions of the dust halo, polarization could not be detected with 3σ confidence.

vectors imply a large-scale organized magnetic structure over distances of about 200 parsecs. As the flow of partially ionized molecular gas and dust is easier along field lines than across them, this may assist in channelling clouds towards the starburst nucleus.

In contrast, outside the torus the vectors (shown in white) have very varied orientations, but mainly lie parallel to the outer flux contours of the dusty halo. This bubble-type field has not previously been seen in any other galaxy^{15,16}, but is reminiscent of magnetic structures seen in the solar neighbourhood¹⁷. In both our Galaxy¹⁷ and M82, the expansion of magnetic bubbles could be driven by kiloparsec-scale winds.

Because the submillimetre dust emission is optically thin, this polarization technique is sensitive to all magnetically aligned grains along a line of sight. The percentage polarization (depicted as proportional to vector length in Fig. 1) is not a measure of magnetic field strength, but can be a diagnostic of grain characteristics¹⁸. In M82, polarization is detected nearly everywhere that there is significant dust emission, with a median level of 2.8%—this is similar to 850- μm polarization values seen in massive star-forming regions in our own Galaxy, such as W3 (median of 2.5%, JCMT data). Thus, there is no reason to infer very different grain properties in M82, and the magnetic properties of the regions of high-mass star-formation may be similar.

We conclude that the major magnetic features found in M82 are ordered fields over scales of hundreds of parsecs within the torus or bar, and an outer bubble-like field associated with the dusty halo, with a diameter of at least 1 arcmin, or 1 kpc. The field orientations in the dust associated with active star formation are not in the galactic plane, nor poloidal within the wind from the nucleus. Both of the latter morphologies have been seen in near-infrared polarimetry¹¹ of M82, but that technique is more sensitive to large column densities of cold dust through the galactic plane as a whole. The phenomena seen in submillimetre polarimetry, even in this one galaxy, have thus given unique insights into physical processes associated with the starburst that cannot be studied at other wavelengths. $\square$

Received 15 November 1999; accepted 15 February 2000.

- Hough, J. H. in *Polarimetry of the Interstellar Medium* (eds Roberge, W. G. & Whittet, D. C. B.) 569–586 (ASP Conference Series Vol 97, 1996).
- Hildebrand, R. H. Magnetic fields and stardust. *Q. J. Astron. Soc.* **29**, 327–351 (1988).
- Dowell, C. D. *et al.* Submillimeter array polarimetry with HERTZ. *Astrophys. J.* **509**, 588–598 (1998).
- Greaves, J. S., Holland, W. S., Minchin, N. R., Murray, A. G. & Stevens, J. A. Submillimetre polarization and constraints on dust grain alignment. *Astron. Astrophys.* **344**, 668–674 (1998).
- Wilson, C. D. *et al.* Submillimeter continuum emission in the ρ Ophiuchi molecular cloud: filaments, arcs and an unidentified far-infrared object. *Astrophys. J.* **513**, L139–L142 (1999).
- Hughes, D. H., Gear, W. K. & Robson, E. I. The submillimetre structure of the starburst nucleus in M82—a diffraction-limited 450-micron map. *Mon. Not. R. Astron. Soc.* **270**, 641–649 (1994).
- Neininger, N., Guelin, M., Keln, U., Garcia-Burillo, S. & Wielebinski, R. ^{13}CO at the centre of M82. *Astron. Astrophys.* **339**, 737–744 (1998).
- O’Connell, R. W., Gallagher, J. S. II, Hunter, D. A. & Colley, W. N. Hubble Space Telescope imaging of super star clusters in M82. *Astrophys. J.* **446**, L1–L4 (1995).
- Alton, P. B., Davies, J. I. & Bianchi, S. Dust outflows from starburst galaxies. *Astron. Astrophys.* **343**, 51–63 (1999).
- Reuter, H.-P., Klein, U., Lesch, H., Wielebinski, R. & Kronberg, P. P. The magnetic field in the halo of M82—Polarized radio emission at 6.2 and 3.6 cm. *Astron. Astrophys.* **282**, 724–730 (1994); erratum in *Astron. Astrophys.* **293**, 287–288 (1995).
- Jones, T. J. The magnetic field geometry in M82 and Cen A. *Am. Astron. Soc. Meeting* **193**, No. 76.03 (American Institute of Physics, Woodbury, New York, 1998).
- Holland, W. S. *et al.* SCUBA: a common-user submillimetre camera operating on the James Clerk Maxwell Telescope. *Mon. Not. R. Astron. Soc.* **303**, 659–672 (1999).
- Greaves, J. S., Jenness, T., Chrysostomou, A. C., Holland, W. S. & Berry, D. S. in *Imaging At Radio Through Submillimeter Wavelengths* (ed. Mangum, J.) (ASP Conference Series, in the press).
- Murray, A. G., Ade, P. A. R. & Griffin, M. J. in *Polarimetry of the Interstellar Medium* (eds Roberge, W. G. & Whittet, D. C. B.) 116–117 (ASP Conference Series Vol. 97, 1996).
- Beck, R., Brandenburg, A., Moss, D., Shurukov, A. & Sokoloff, D. Galactic magnetism: recent developments and perspectives. *Annu. Rev. Astron. Astrophys.* **34**, 155–206 (1996).
- Lesch, H. & Harnett, J. Galactic dynamics and magnetic fields. *Astron. Astrophys.* **268**, 58–64 (1993).
- Vallee, J. P. Observations of the magnetic fields inside and outside the Milky Way, starting with globules (~ 1 parsec), filaments, clouds, superbubbles, spiral arms, galaxies, superclusters, and ending with the cosmological universe’s background surface (at ~ 8 Teraparsecs). *Fundamentals Cosmic Phys.* **19**, 1–89 (1997).
- Hildebrand, R. H. & Dragovan, M. The shapes and alignment properties of interstellar dust grains. *Astrophys. J.* **450**, 663–666 (1995).

Acknowledgements

We thank T. Jones for discussions, and D. Berry and A. Chrysostomou for work on the data reduction software. The JCMT is operated by the Joint Astronomy Centre on behalf of the UK Particle Physics and Astronomy Research Council, the Netherlands Organisation for Pure Research, and the National Research Council of Canada.

Correspondence and requests for materials should be addressed to J.S.G. (e-mail: j.greaves@jach.hawaii.edu).

Mechanisms of extensive spatiotemporal chaos in Rayleigh–Bénard convection

David A. Egolf^{*†}, Ilarion V. Melnikov^{*‡}, Werner Pesch[§] & Robert E. Ecke^{*†}

^{*} Center for Nonlinear Studies, Theoretical Division, [†] Condensed Matter and Thermal Physics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[§] Theoretical Physics II, Physikalisches Institut, University of Bayreuth, Universitätstrasse 30 D-95440, Bayreuth, Germany

Spatially extended dynamical systems exhibit complex behaviour in both space and time—spatiotemporal chaos^{1,2}. Analysis of dynamical quantities (such as fractal dimensions and Lyapunov exponents³) has provided insights into low-dimensional systems; but it has proven more difficult to understand spatiotemporal chaos in high-dimensional systems, despite abundant data describing its statistical properties^{1,4,5}. Initial attempts have been made to extend the dynamical approach to higher-dimensional systems, demonstrating numerically that the spatiotemporal chaos in several simple models is extensive^{6–8} (the number of dynamical degrees of freedom scales with the system volume). Here we report a computational investigation of a phenomenon found in nature, ‘spiral defect’ chaos^{5,9} in Rayleigh–Bénard convection, in which we find that the spatiotemporal chaos in this state is extensive and characterized by about a hundred dynamical degrees of freedom. By studying the detailed space–time evolution of the dynamical degrees of freedom, we find that the mechanism for the generation of chaotic disorder is spatially and temporally localized to events associated with the creation and annihilation of defects.

The strong irregularity in both space and time observed in Rayleigh–Bénard convection is found in other complex dynamical systems such as heart tissue, nonlinear chemical and fluid patterns, planetary atmospheres and oceans, fluid turbulence, and environmental ecosystems. In some instances, the effective degrees of freedom in the problem (roughly, the average number of independent variables needed to describe the system) are thought to be extensive^{6–8}. In others, new behaviours are found at all scales. In both cases, identifying the dynamical degrees of freedom and the instability mechanisms leading to disorder is of great importance. For example, better prediction of the nucleation and paths of tornadoes or hurricanes, the evolution of the global climate, or the transition to cardiac arrhythmia might be expected. Although intuitive prediction can often be useful, a general quantitative technique for identifying instability mechanisms that lead to loss of predictability in physically realizable, complex, spatially extended systems has not yet been achieved. For developing and testing such methods, pattern formation in nonlinear, nonequilibrium systems

[‡] Present address: Department of Physics, Duke University, Durham, North Carolina 27708, USA.

is ideal because there are many high-precision laboratory experiments on pattern dynamics^{1,5} and because numerical simulation of the evolution equations for such systems is now within the grasp of present-day supercomputers.

Some of the most detailed experiments on complex dynamics in pattern-forming systems have studied Rayleigh–Bénard convection (RBC). An RBC experiment consists of a fluid layer of depth d confined between two horizontal conducting plates held at temperatures T_{top} and T_{bottom} . For temperature differences $\Delta T = T_{\text{bottom}} - T_{\text{top}}$ less than a critical temperature difference ΔT_c , the fluid is stationary. When $\Delta T > \Delta T_c$, the stationary state is unstable and the fluid begins to flow⁵ with the ‘distance’ from the onset of flow expressed as $\epsilon = \Delta T/\Delta T_c - 1$. For systems with small aspect ratios (lateral dimension L relative to d), the flow is typically organized into parallel convective ‘rolls’ of diameter approximately equal to d , with each roll consisting of hot fluid moving upward on one side and cold fluid moving downward on the other. The convective flow carries heat and enhances heat transport, making convection an important process in the atmosphere and oceans and in many technological applications. For large aspect ratios (L/d of order 100), Morris *et al.*⁹ discovered the persistent state of spiral defect chaos (SDC), consisting of locally parallel rolls arranged in temporally chaotic, spatially complex patterns. This work led to many more experiments, yielding a large quantity of high-precision data⁵. In these experiments, the complex two-dimensional convective patterns are visualized using the optical shadowgraph technique¹⁰. Virtually indistinguishable patterns (see, for example, Fig. 1) are observed in simulations of the Boussinesq equations, the standard hydrodynamic equations describing convection¹¹. The patterns have been characterized by global Fourier transform methods^{9,12–14}, with local roll properties such as curvature and wavenumber variations^{13–15}, and by the statistics of spiral-defect populations^{15,16}. Researchers have also studied the mechanism of spiral creation and stability both experimentally and theoretically^{17–19}. However, the direct dynamical degrees of freedom are still unknown.

A common method for quantifying the dynamics of nonlinear systems is to study the evolution of states that are nearly identical to a reference state. The divergences of these states from the reference state are characterized by a set of exponents called the Lyapunov spectrum $\lambda_1 > \lambda_2 > \dots$ (ref. 3). Each Lyapunov exponent λ_i quantifies the average exponential rate of growth of the perturbation represented by the corresponding Lyapunov vector, which is a particular perturbation (out of all of the possible choices of perturbations) away from the reference state, with each Lyapunov vector orthogonal to all others. For RBC, the perturbations represented by the Lyapunov vectors are small differences in the



Figure 1 Spiral defect chaos in Rayleigh–Bénard convection. The instantaneous mid-plane temperature field is shown for $\epsilon = 0.8$.

temperature and/or velocity fields with respect to a reference state exhibiting SDC. The evolution of the reference state is computed using the Boussinesq equations, and the evolutions of the Lyapunov vectors are computed according to the linearization of the Boussinesq equations about the time-dependent reference state. (The linearized equations ensure that the exponents describe the divergence of states that differ only infinitesimally from the reference.) The ‘direction’ of each Lyapunov vector characterizes the spatial distribution of perturbations in the temperature and velocity fields (with each component of the vector representing the perturbation at a particular point in space for the temperature or velocity fields).

The (fractal) Lyapunov dimension D is a measure of the complexity of the dynamics (that is, how chaotic the system is) and can be calculated by the Kaplan–Yorke formula, which expresses D in terms of the Lyapunov exponents λ_i . From arguments first given by Ruelle⁶, we expect that in large systems distant regions will be dynamically independent, so D will grow in proportion to the volume L^{d_s} , where d_s is the spatial dimensionality ($d_s = 2$ for SDC) and L is the lateral length scale; that is, the dimension is extensive with intensive dimension density $\delta = D/L^{d_s}$. This chaotic extensivity can be phrased in stronger terms by studying the Lyapunov spectral density $\lambda(i/L^{d_s})$ as $L \rightarrow \infty$ (where i is the same index as in λ_i). For extensively chaotic systems $\lambda(i/L^{d_s})$ becomes invariant. Even for relatively low-dimensional dynamical systems it is extremely difficult to measure enough Lyapunov exponents from experimental data to compute D . Thus, researchers have instead used relatively simple spatially extended model systems such as the Kuramoto–Sivashinsky equation⁷ and the complex Ginzburg–Landau (CGL) equation⁸, for which simple but computationally intensive methods have allowed the computation of the Lyapunov spectra for many different system sizes L . For these systems, the chaotic dynamics was shown to be extensive.

To test whether an actual experimental system exhibits extensive chaos, we studied the Boussinesq equations computationally. Even though several simple models exhibit extensive chaos, it is not obvious a priori that the chaos in RBC is extensive. RBC differs from

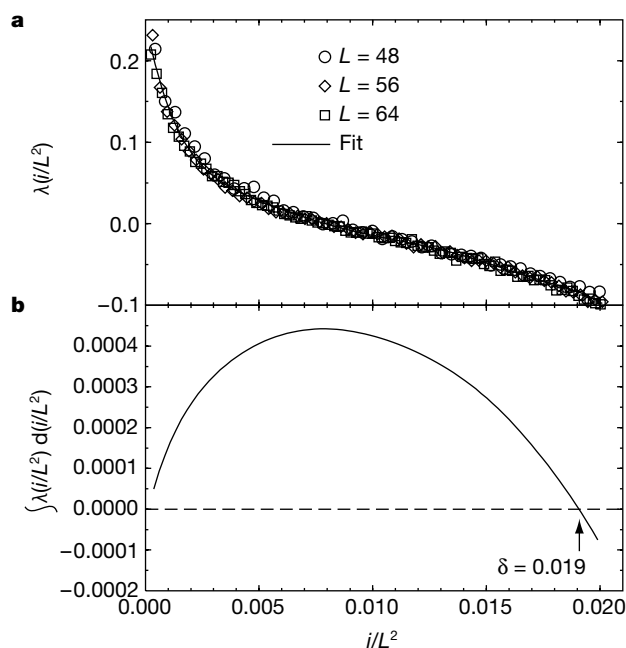


Figure 2 Lyapunov spectral density and its integral. **a**, $\lambda(i/L^2)$ versus system-size-normalized index i/L^2 for $\epsilon = 0.8$ and $L = 48$ (circles), $L = 56$ (diamonds) and $L = 64$ (squares). The solid line is a seventh-degree polynomial fit to the data. **b**, Integral of $\lambda(i/L^2)$ with respect to i/L^2 versus i/L^2 computed from the curve fit in **a**. The dimension density $\delta \approx 0.019$ is determined by the zero intercept of the data.

the simple models in significant ways. In RBC, pressure modes that act instantaneously and decay slowly in space could couple distant spatial regions. Also, simple non-chaotic RBC states can exist for the same experimental parameters at which spatiotemporal chaos is found²⁰.

We integrated the Boussinesq equations using a pseudospectral method with time splitting of the operators. To minimize boundary effects and emphasize bulk properties, periodic boundary conditions were employed. Our code is consistent with exact linear and weakly nonlinear results. We verified numerical convergence with respect to spatial and temporal resolutions and integration times. The Lyapunov spectrum was obtained by simultaneously evolving up to $N = 128$ orthogonal fields (the Lyapunov vectors) according to the linearization of the Boussinesq equations about a particular solution exhibiting SDC. The N Lyapunov vectors are the N (orthogonal) perturbations with the largest average growth rates λ_i . Further details will be provided elsewhere (D.A.E. *et al.*, manuscript in preparation).

Figure 2a shows the Lyapunov spectral density $\lambda(i/L^2)$ for the SDC state of RBC at $\epsilon = 0.8$. The Lyapunov spectral densities for system sizes $L = 48, 56$ and 64 collapse onto a single curve, demonstrating unambiguously that SDC is extensive. These data represent the first (to our knowledge) direct demonstration of the principle of extensivity of the dynamics of an experimentally relevant spatiotemporal chaotic system. The value of the dimension density δ , calculated using the Kaplan–Yorke formula, is the value of i/L^2 such that the area under the curve shown in Fig. 2a from 0 to $\delta = i/L^2$ is zero. This procedure is shown schematically in Fig. 2b and yields $\delta = 0.019$ (in units of d^{-2}) corresponding to a fractal dimension $D \approx 80$ for $L = 64$. We note that the number of dynamical degrees of freedom ($\sim 10^2$) is smaller by several orders of

magnitude than the number of spatial modes obtained by orthogonal decomposition ($\sim 10^4$, using either Karhunen–Loeve²¹ or Fourier methods). It is the same order of magnitude, however, as estimates¹² from experiments ($D \approx 200$) based on measured correlation lengths and adjusted for our system size.

Because the Lyapunov spectrum is an infinite-time average, the intriguing dynamical behaviour of the underlying Lyapunov vectors is often overlooked. The spatial distributions of the perturbations, as well as their growth rates, vary substantially in time. The growth rates are quantified by the time-dependent finite-time Lyapunov spectrum of exponents $\lambda_i^{(\Delta\tau)}(t)$ describing the exponential divergence of similar states over the interval $(t, t + \Delta\tau)$. The time-dependent Lyapunov vector directions (spatial distributions of perturbations) and finite-time Lyapunov exponents reflect the dynamical features within the evolution of the system itself^{22,23}. Here we study the dynamical degrees of freedom in greater detail by comparing the space–time structures of the Lyapunov vectors to the evolution of the system in order to reveal the mechanism of the spatiotemporal chaos in RBC.

For our system, the positive Lyapunov exponents correspond to highly localized vectors, whereas the zero and negative exponents correspond to vectors that are more homogeneously distributed in space. The localization of the Lyapunov vectors with positive exponents is correlated with regions of the pattern that are linearly unstable as a result of the rolls being too narrow or too wide^{15,24}. The finite-time Lyapunov exponents have considerable fluctuations with the largest positive Lyapunov exponents having the greatest standard deviations. The highest maxima in the time sequences of finite-time exponents correspond to defect nucleation events. Figure 3a shows the magnitudes of the temperature–field components of the Lyapunov vector corresponding to the largest Lyapunov

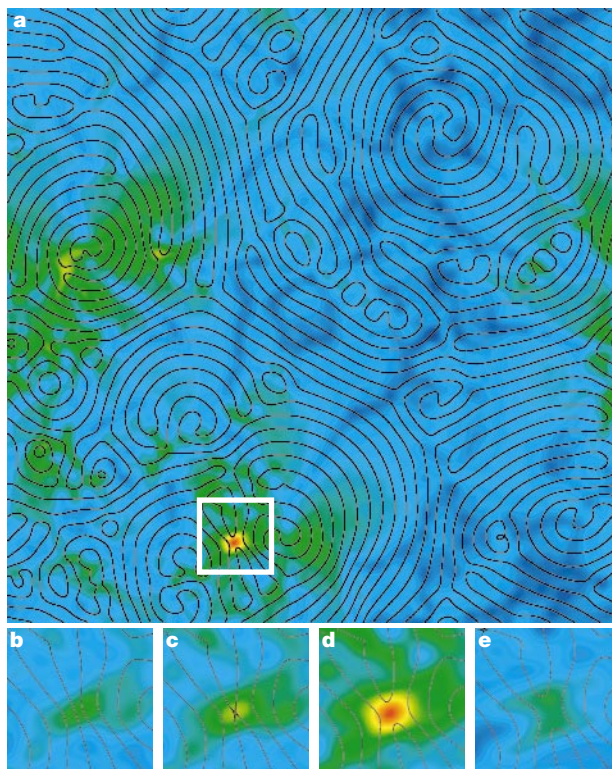


Figure 3 Magnitudes of the temperature–field components of the Lyapunov vector corresponding to the largest Lyapunov exponent, scaled by the first finite-time exponents. The data are colour-coded by magnitude, and superimposed on the underlying roll pattern. **a**, Whole 64×64 pattern. **b**, Blown-up section $\Delta t = 0.3750\tau_v$ before **a** (the timescale τ_v , the vertical thermal diffusion time, is roughly the timescale of the fastest

events within the dynamics, such as the defect nucleation). **c**, Blown-up section $\Delta t = 0.1875\tau_v$ before **a**. **d**, Blown-up section of **a**. **e**, Blown-up section $\Delta t = 0.1875\tau_v$ after **a**. The colour scale (blue–cyan–green–yellow–red) is linear in the $1/4$ power of the magnitude of the scaled vector. The size of the inset box is about one chaotic domain $\xi_0^2 = L^2/D$.

exponent, suitably scaled and colour-coded to represent local magnitude (red is highest, blue is smallest). Much of the pattern is quiescent but there is a strong event in the lower left. This localization is common for the vectors corresponding to positive Lyapunov exponents ('chaotic') and is further evidence for the Ruelle picture of extensive chaos—a doubling of the system size results in new Lyapunov vectors localized to the newly added area of the system.

Figure 3b–e shows an expanded, time-ordered view of a particular event in Fig. 3a. The largest finite-time Lyapunov exponent quickly builds (Fig. 3b and c) to a maximum (Fig. 3d) at the moment of roll-breaking, followed by a rapid relaxation (Fig. 3e) to a negative value. The Lyapunov vector is strongly localized to the small spatial region at which the breakage occurs, signifying that the evolution of the system is highly susceptible to small perturbations in this particular region, with minute details of the state determining the exact breaking point and defect motions and ultimately the future dynamics of the system. (In contrast, small differences between states in regions away from roll breaking/reconnection events decay exponentially.) Shortly after the rolls break/reconnect, the dynamics of this Lyapunov vector is mostly relaxational as the defects diffuse away. This scenario of strong localization of the Lyapunov vector and positive spikes in the finite-time Lyapunov exponent correlated with roll breaking/reconnection events is intermittently repeated throughout the chaotic evolution of the system. (The strongly chaotic events include not only defect-generating skewed varicose events shown in Fig. 3, but also cross-roll events in regions of large local wavelength²⁴.)

These data demonstrate that the mechanism generating the chaotic dynamics in the SDC state of RBC is roll breaking/reconnection rather than the more apparent motion of the large spiral defects (in direct contrast to the active cores of chemical spirals). This mechanism for the generation of chaotic disorder makes intuitive sense because major roll reorientations occur through roll breaking/reconnection. Thus, the disorder in the patterns arises from the complicated regions between the spirals rather than in the spirals themselves (although the spirals may contribute to the generation of the complicated regions).

With sufficient computer power, the analysis technique described here is applicable to many complex problems for which the evolution equations are well-known or well-approximated. Although no methods for obtaining Lyapunov vectors from experimental data currently exist, this type of analysis might soon be possible using an experimental technique to repeatedly recreate particular chaotic states or an analysis technique based on recurring substates. □

Received 15 December 1999; accepted 21 February 2000.

1. Cross, M. C. & Hohenberg, P. C. Pattern-formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851–1112 (1993).
2. Cross, M. C. & Hohenberg, P. C. Spatiotemporal chaos. *Science* **263**, 1569–1570 (1994).
3. Eckmann, J.-P. & Ruelle, D. Ergodic-theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985).
4. Gollub, J. P. Order and disorder in fluid motion. *Proc. Natl Acad. Sci. USA* **92**, 6705–6711 (1995).
5. Bodenschatz, E., Pesch, W. & Ahlers, G. Recent developments in Rayleigh–Bénard convection. *Annu. Rev. Fluid Mech.* **32**, 709–778 (2000).
6. Ruelle, D. Large volume limit of the distribution of characteristic exponents in turbulence. *Commun. Math. Phys.* **87**, 287–302 (1982).
7. Manneville, P. *Liapunov Exponents for the Kuramoto-Sivashinsky Model* Vol. 280, 319–326 (Lecture Notes in Physics, Springer, Berlin, 1994).
8. Ego, D. A. & Greenside, H. S. Relation between fractal dimension and spatial correlation length for extensive chaos. *Nature* **369**, 129–131 (1994).
9. Morris, S. W., Bodenschatz, E., Cannell, D. S. & Ahlers, G. Spiral defect chaos in large aspect ratio Rayleigh–Bénard convection. *Phys. Rev. Lett.* **71**, 2026–2029 (1993).
10. de Bruyn, J. et al. Apparatus for the study of Rayleigh–Bénard convection in gases under pressure. *Rev. Sci. Instrum.* **67**, 2043–2067 (1996).
11. Decker, W., Pesch, W. & Weber, A. Spiral defect chaos in Rayleigh–Bénard convection. *Phys. Rev. Lett.* **73**, 648–651 (1994).
12. Morris, S. W., Bodenschatz, E., Cannell, D. S. & Ahlers, G. The spatiotemporal structure of spiral-defect chaos. *Physica D* **97**, 164–179 (1996).
13. Hu, Y., Ecke, R. E. & Ahlers, G. Transition to spiral-defect chaos in low Prandtl number convection. *Phys. Rev. Lett.* **74**, 391–394 (1995).
14. Hu, Y., Ecke, R. E. & Ahlers, G. Convection for Prandtl numbers near 1: Dynamics of textured patterns. *Phys. Rev. E* **51**, 3263–3279 (1995).

15. Ego, D. A., Melnikov, I. V. & Bodenschatz, E. Importance of local pattern properties in spiral defect chaos. *Phys. Rev. Lett.* **80**, 3228–3231 (1998).
16. Ecke, R. E., Hu, Y., Mainieri, R. & Ahlers, G. Excitation of spirals and chiral-symmetry breaking in Rayleigh–Bénard convection. *Science* **269**, 1704–1707 (1995).
17. Assenheimer, M. & Steinberg, V. Transition between spiral and target states in Rayleigh–Bénard convection. *Nature* **367**, 345–347 (1994).
18. Cross, M. C. & Tu, Y. Defect dynamics for spiral chaos in Rayleigh–Bénard convection. *Phys. Rev. Lett.* **75**, 834–837 (1995).
19. Plapp, B. B., Ego, D. A., Bodenschatz, E. & Pesch, W. Dynamics and selection of giant spirals in Rayleigh–Bénard convection. *Phys. Rev. Lett.* **81**, 5334–5337 (1998).
20. Cakmur, R. V., Ego, D. A., Plapp, B. B. & Bodenschatz, E. Bistability and competition of spatiotemporal chaotic and fixed point attractors in Rayleigh–Bénard convection. *Phys. Rev. Lett.* **79**, 1853–1856 (1997).
21. Zoldi, S. M., Liu, J., Bajaj, K. M. S., Greenside, H. S. & Ahlers, G. Extensive scaling and nonuniformity of the Karhunen–Loève decomposition for the spiral-defect chaos state. *Phys. Rev. E* **58**, R6903–R6906 (1998).
22. Chaté, H. Lyapunov analysis of spatiotemporal intermittency. *Europhys. Lett.* **21**, 419–425 (1993).
23. Ego, D. A. Dynamical dimension of defects in spatiotemporal chaos. *Phys. Rev. Lett.* **81**, 4120–4123 (1998).
24. Busse, F. H. & Clever, R. M. Instabilities of convection rolls in a fluid of moderate Prandtl number. *J. Fluid Mech.* **91**, 319–335 (1979).

Acknowledgements

We acknowledge helpful discussions with E. Bodenschatz and R. Mainieri. This work was funded by the US Department of Energy, and significant computational resources were provided on the Nirvana machines of the Advanced Computing Laboratory at Los Alamos National Laboratory. The development of a related version of the Boussinesq equation solver benefited from the support of a US National Science Foundation grant from the Division of Materials Research.

Correspondence and requests for materials should be addressed to D.A.E. (e-mail: egolf@lanl.gov).

Phase transitions in the incoherent lattice fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

R. P. Sharma*, S. B. Ogale*, Z. H. Zhang†, J. R. Liu†, W. K. Chu†, Boyed Veal‡, A. Paulikas‡, H. Zheng‡ & T. Venkatesan*

* Center for Superconductivity Research, Department of Physics, University of Maryland, College Park, Maryland 20742, USA

† Texas Centre for Superconductivity, University of Houston, Houston, Texas 77204, USA

‡ Materials Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

The growing body of experimental evidence^{1–7} for the existence of complex textures of charges and spins in the high-temperature superconductors has drawn attention to the so-called 'stripe-phase' model^{8–12} as a possible basis for the mechanism of superconductivity in these materials. Such observations have until now been restricted to systems where the texture dynamics are slow or suppressed altogether, and do not include the important case of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. It seems likely that the dynamic behaviour of stripes^{12,13}, which has been suggested to undergo several phase transitions as a function of temperature¹², should also be reflected in the lattice properties of the host materials, and this forms the motivation for our present experiments. Specifically, we use MeV helium ion channelling, an ultrafast real-space probe of atomic displacements (with sub-picometre resolution), to probe incoherent lattice fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ as a function of temperature and oxygen doping. We detect lattice fluctuations that are larger than the expected thermal vibration component, and which show anomalies characteristic of the phase transitions anticipated for a dynamic stripe phase. Comparison of our lattice results with single-particle-tunnelling and photoemission data highlights the

importance of spin–charge separation phenomena in the copper oxide superconductors.

The close proximity of the insulating antiferromagnetic and superconducting states in high- T_c copper oxides is crucial to the mechanism of high- T_c superconductivity. In this context, other workers^{8–11} have addressed the issue of the behaviour of doped holes in an antiferromagnetic background. Their pioneering studies revealed that such a system has a tendency to phase separate, with the doped holes forming stripes that separate hole-free antiferromagnetic insulating phase domains. These self-organized complex textures of charges and spins introduce intermediate length, time and temperature scales in the physics of the problem. Such a “stripe phase” is presumably dynamic¹³, and undergoes several phase crossovers or transitions as a function of temperature¹². The doping-temperature phase diagram for such systems has been addressed^{14,15} on phenomenological grounds based on a host of experimental data^{14,16–18}. There is considerable evidence for the existence of metallic stripes in the La_2CuO_4 family of superconductors: ordered stripes as in $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$ (ref. 1), dynamically fluctuating stripes as in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (refs 1 and 2), or pinned and meandering stripes as in lightly doped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (ref. 3), presumably obtainable because of very slow stripe fluctuations in this system. There is recent evidence⁷ for one-dimensional charge transport in $\text{La}_{2-x-y}\text{Nd}_y\text{Sr}_x\text{CuO}_4$, demonstrating that stripe order readily develops in the normal state of lanthanum copper oxides. However, the corresponding evidence is less direct in other high- T_c materials such as YBCO ($\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$) and BSCCO ($\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$). Neutron-scattering data^{4,5} on underdoped YBCO and ARPES data on BSCCO (ref. 4) suggest that there are stripe-like but more disordered local charge fluctuations in these superconductors. If brisk dynamics of stripes is a reason for the apparent disorder, ultrafast probes should be used to investigate such systems.

The search for the possible signatures of dynamic stripes can be based upon the hints obtained from different experiments. A study¹⁹ of Zn-doped YBCO by optical reflectivity concludes that a coupling of phonons to spin fluctuations accompanies the opening of the spin gap. In a different system (CaV_4O_9), opening of the spin gap is found²⁰ to be associated with strong temperature-dependent lattice distortions. Recent polarized X-ray scattering experiments²¹ in the 214 copper oxides have identified the presence of a variety of phases such as the low-temperature orthorhombic (LTO) and low-temperature tetragonal (LTT) phases; a specific proposal was made where the antiferromagnetic region had the LTT phase and the metallic stripes had the LTO phase. The lattice parameters of the two phases differ by as much as 0.01 Å. We would therefore expect a lattice fluctuation at the phase boundary between the stripe and

antiferromagnetic phases (ref. 22 claims that in a simple t–J model the striped states of the system are not ground states, whereas if one includes the coupling of the electrons with lattice distortions then stripe phases may be allowed), provided the fluctuation frequencies are comparable or slower than phonon frequencies for the lattice to follow. The spin-fluctuation frequencies in the 214 phase have been estimated²³ and have shown a strong temperature dependence with the spin-fluctuation frequencies changing from 10^{13} Hz at room temperature to 10^{10} Hz at low temperatures. Hence, experiments with the probe response time faster than 10^{-13} seconds would then be able to follow these fluctuations. One such experiment is MeV ion channelling spectrometry, where the scattering times are $<10^{-17}$ seconds. On this timescale, most dynamic condensed matter processes would appear static.

Ion channelling²⁴ is a unique method, providing a direct real space probe of extremely small (sub-picometre) uncorrelated displacements (static and dynamic) of atoms in single crystalline materials. Ion channelling occurs when energetic ions, incident along a major crystallographic direction, are steered by a series of glancing collisions with the atoms of close packed atomic rows or planes. The critical angle for channelling to occur, or the full-width-at-half-maximum (FWHM) of the channelling angular scan, depends on the incident ion energy, the atomic numbers of the projectile and target, the interatomic spacing, electron screening potential and most importantly, any displacements (dynamic or static) of the atoms from their regular lattice sites. The FWHM is strongly affected by incoherent atomic displacements (both static and dynamic), as opposed to an equivalent coherent change in lattice parameter (for example, thermal expansion). The magnitude of the atomic displacement u can be extracted from the measured FWHM using Lindhard’s continuum model²⁵, with appropriate corrections based upon a simulation²⁶, as we do here.

Single crystals of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ were grown by a partially non-stoichiometric melting technique at Argonne National Laboratory. These were annealed in flowing oxygen at 450 °C to 500 °C for about 72 hours and showed a sharp ($\Delta T_c < 1$ K) superconducting transition at 92.5 K. Different sets of such crystals were annealed a second time for about 180 hours in the requisite O_2/N_2 ambient to reduce the O doping to get crystals with T_c at 65 K and 45 K respectively, and also a non-superconducting crystal. The ion channelling measurements were carried out using a well collimated beam (0.5 mm diameter and $< 0.01^\circ$ divergence) of 1.5 MeV He^+ ions obtained from a 1.7 MV tandem pelletron accelerator²⁷. The crystal was suitably mounted on a precision five-axis goniometer with an angular resolution less than 0.01° and could be cooled down to 33 K with the help of a closed-cycle refrigeration unit. Necessary precautions were taken to keep the effective pressure of condensable

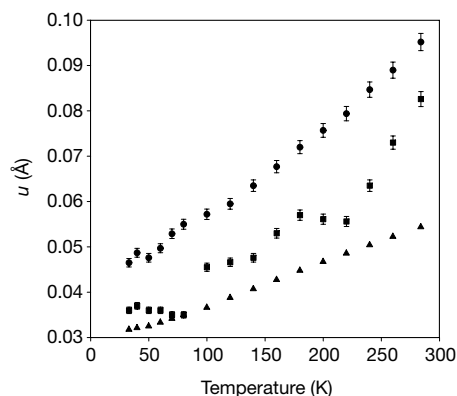


Figure 1 Temperature dependence of atomic displacements for $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ($x \approx 0.05$ and 0.065). The figure shows the atomic displacement u as obtained from the FWHM of channelling angular scans in a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ crystal ($T_c = 92.5$ K; squares) and a non-

superconducting crystal (circles) as a function of temperature. The Debye behaviour normalized at the 80 K value of u for the 92.5 K sample is also shown (triangles).

gases to a negligible level ($\sim 3 \times 10^{-11}$ torr) at the target surface; the background pressure in the target chamber was maintained at $\sim 2 \times 10^{-8}$ torr.

The backscattered He^+ particles were analysed using an annular silicon surface barrier detector of 300 mm^2 active area with a central hole 4 mm in diameter which was mounted along the beam axis at a distance of about 5 cm from the target. This arrangement provided good statistics ($\sim 10^4$ counts in the gate at each angular setting in random direction) for back scattered (170°) particles at a dose of only 3 nC of the incident He-ions in a channelling angular scan. The [001] single crystals of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ with suitable oxygen doping were first aligned precisely parallel to the incident beam direction. Angular channelling scans were made by measuring the Rutherford back scattering (RBS) yield as a function of the tilt angle about the [001] axis of the sample at 16 different temperatures in the range 300 K to 30 K in each case respectively. In each case two to three sets of measurements were made using eight to ten high quality YBCO single crystals of respective T_c values. A standard deviation of less than 1% was determined in these measurements.

The u versus temperature data for the optimally oxygen-doped and underdoped non-superconducting YBCO samples are shown in Fig. 1. It is interesting to note that the data for the optimally doped sample show well-defined structures in the temperature dependence of the u value, which are considerably suppressed in the case of the non-superconducting sample. At least three crossover temperatures can be visually identified. To bring out these crossover or transition temperatures more clearly we subtracted the thermal vibration component, represented by the reference Debye curve (Fig. 1) from the u values for the optimally-doped ($T_c = 92.5$ K) and underdoped ($T_c = 65$ K, 45 K) samples. The Debye curve was obtained with the constraint that the excess distortion measured

above the thermal background cannot be negative. The 'excess' lattice distortions (u_{ex}) obtained by subtraction are shown in Fig. 2.

Let us first consider the optimally doped case. From room temperature to a temperature of about 230 K there is a rapid drop in u_{ex} . At about 230 K (T_1) the nature of u_{ex} is indicative of some kind of phase transition in the system which enhances these fluctuation effects (dynamic or static). One more cusp is seen close to 150 K (T_2). Near the electrically measured superconducting transition a sudden drop is seen in u_{ex} at T_3 . Below T_3 , u_{ex} increases with decreasing temperature. These features are also reproduced fairly consistently in the data for the $T_c = 65$ K and 45 K oxygen-underdoped samples, with the drop at T_3 scaling with the T_c , and below T_c an increasing u_{ex} with decreasing temperature. The data about the characteristic transition or crossover temperatures as a function of oxygen doping are summarized in Fig. 3.

One of the features that emerges from our channelling data is the non-analytic behaviour of u_{ex} in the normal state near T_1 and T_2 . These occur exclusively in the incoherent lattice displacements, when none at all is known to occur in the lattice structure. We seek the origin of the anomalous lattice fluctuations in the influence of the electron system on the lattice by the known coupling between the lattice fluctuations and the low energy excitations of the electron system. However, if the metallic state is regarded as a Fermi liquid, the strength of this influence should be very small in these systems, in view of an expected large Fermi energy (~ 1 eV) and thereby a small value of Migdal parameter ($\hbar\omega_D/E_F < 0.035$). Even if we push the Fermi-liquid theory to its limit and obtain anomalous lattice fluctuations, their evolution with temperature would not exhibit sudden changes such as those observed in our experiment, because the theory does not allow the required cooperativity of interactions and related phenomena for these to occur. Our data thus emphasize that the normal state in YBCO cannot be understood in terms of the conventional theory of metals.

From Fig. 1, in the case of the non-superconducting sample ($\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, $x \approx 0.65$), the rate of increase of u is higher than for the Debye u value. This implies a gradual increase of u_{ex} with temperature. This in itself is interesting and should serve as a reference for the analysis of the magnitude and the nature of non-monotonic dependence seen in the superconducting samples. It is generally thought that the non-superconducting samples have a stronger propensity to form stripes, although their dynamics largely reflects a classical character⁷. In such a classical model, the stripe formations at low temperatures should be disrupted gradually with increasing temperature, thereby giving an increasing amplitude of

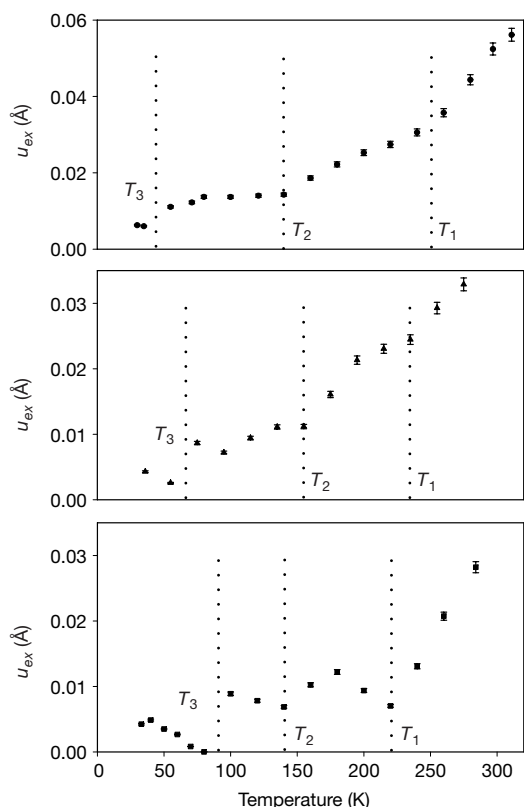


Figure 2 Temperature dependence of excess atomic displacement in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. The figure shows the excess atomic displacement u_{ex} as obtained after subtracting the Debye component from the u values for 45 K (top), 65 K (middle) and 92.5 K (bottom) superconducting YBCO samples as a function of temperature.

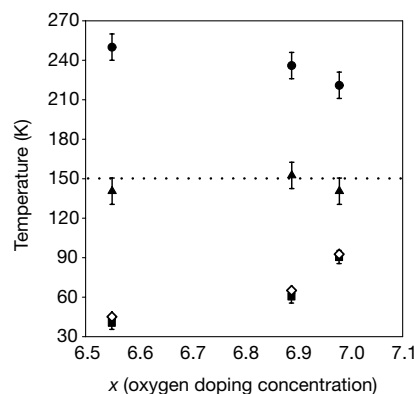


Figure 3 Oxygen-doping dependence of phase transition or crossover temperatures. The figure shows the phase transition or crossover temperatures as indicated in Fig. 2, as a function of oxygen-doping concentration: T_1 (circles); T_2 (triangles); T_3 (squares). T_c is indicated by a white diamond.

the excess lattice fluctuation (easily identified with the charge fluctuation) with increasing T . The inelastic neutron-scattering data show that the density of low-energy spin fluctuations is a monotonically decreasing function of temperature⁵. This emphasizes the difference in the dynamics of spin and charge degrees of freedom in these systems, strongly suggesting a spin-charge separation scenario. In the case of the superconducting samples, one would expect the extra quantum correlations to boost the stripe (charge) fluctuation frequency, thereby bringing down the amplitude of the observed non-Debye (anomalous) response. The role of such correlations and their evolution across the crossovers is thus primarily one of bringing down the excess fluctuations. We thus next address the recent observations²⁸ of remarkable phonon anomalies in $\text{Y}_1\text{Ba}_2\text{Cu}_3\text{O}_{7-x}$, which reveal the presence of charge fluctuations. These anomalies are reflected strongly for a phonon mode representing motions of planar oxygen atoms within the CuO planes, which also clearly influence the channelling of ions moving along the c -axis (perpendicular to CuO planes). The largest phonon anomalies are seen²⁸ in their $\text{Y}_1\text{Ba}_2\text{Cu}_3\text{O}_{6.35}$ sample, which is where we saw the highest amplitude of non-Debye response.

In view of the appearance of two characteristic crossover temperatures T_1 and T_2 in the normal state regime of our data, we seek their possible connection to the temperatures T_1^* and T_2^* in the stripe phase scenario¹². T_1^* is suggested to characterize the aggregation of holes into stripes (the self-organized structure termed topological doping with locally quasi-one-dimensional electronic character) and T_2^* is suggested to correspond to the onset of pairing and superconducting correlations within an isolated stripe.

Identification of T_2^* with the pseudo-gap temperature deduced from experiments¹⁵ or the temperature for pairing within an isolated stripe¹² suggests a nearly fixed value of ~ 150 K for T_2^* over a range of moderate dopant concentrations. This is indeed what we observe for T_2 . However, a weak dependence also of T_1 on oxygen doping differs distinctly from a relatively rapid dependence of T_1^* predicted by the stripe-phase model¹² or of the single-particle excitation energy gap (Δ_p) obtained by Giaever tunnelling and photoemission measurements²⁹. As the lattice fluctuations we studied can be identified with charge fluctuations, whereas the tunnelling and photoemission measurements probe fermions with both their spin and charge complements, the differences in the systematics of the two sets of data suggest that the energy scales associated with the spin and charge quantum numbers are different in the normal state. This once again suggests a non-Fermi-liquid scenario with spin-charge separation.

If one identifies T_1 as the temperature where the stripe phase forms, then one can interpret the sudden increase in the u value with decreasing temperature, based on the fact that the formation of the stripes could lead to two different regions with different lattice constants²¹. As a result, one may expect dechannelling effects owing to both static and dynamic effects, the latter arising from the fluctuation of the metallic-insulating (antiferromagnetic) boundary^{12,13}. With further decrease in temperature, the u value begins to drop until the second transition T_2 is reached. If T_2 is interpreted as the spin-gap and Cooper-pair-formation temperature within a one-dimensional stripe, it implies that enhanced excess lattice fluctuations are produced when the individual stripes become superconducting. This result is consistent with other results^{19,20}, in which lattice instabilities are reported near the spin-gap formation temperature. At T_c a significant drop in the excess fluctuation is seen, suggesting increased coherence of the lattice. Below T_3 or T_c , the excess fluctuations show a small increase with decreasing temperature, which should be related to the spin fluctuation dynamics. If the metallic-insulating (antiferromagnetic) boundary is fluctuating^{12,13} and is reflected in the spin fluctuation measurement, then on the basis of the 214 data of ref. 23, the boundary fluctuation slows down with decreasing temperature. This means that the ability of the lattice to follow

the spin fluctuation is enhanced resulting in larger measurable lattice fluctuations.

In conclusion, the incoherent lattice fluctuations in YBCO show a non-analytic behaviour as a function of temperature at two characteristic temperatures (T_1 and T_2) well above T_c , as well as at T_c . The weak dependence of T_2 on oxygen doping conforms to the suggested behaviour of the pseudo-gap formation temperature, or the pairing temperature within a single stripe. However, a weak dependence also of T_1 on oxygen doping differs from a relatively rapid dependence of the single-particle excitation gap: a clear difference in the spin and charge dynamics. The observed sequence of transitions at T_1 , T_2 and T_c appears to be consistent with the stripe-phase model, wherein the temperature sequence of stripe formation, spin-gap and pair formation, and the superconducting transition is suggested. The increasing excess lattice fluctuation below T_c is an indication of slowing boundary fluctuations, which the lattice could follow more easily. □

Received 14 September 1999; accepted 14 February 2000.

- Tranquada, J. M. *et al.* Neutron-scattering study of stripe-phase order of holes and spins in $\text{La}_{1.48}\text{Nd}_{0.48}\text{Sr}_{0.12}\text{CuO}_4$. *Phys. Rev. B* **54**, 7489–7499 (1996).
- Cheong, S.-W. *et al.* Incommensurate magnetic fluctuations in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **67**, 1791–1794 (1991).
- Borsa, F. *et al.* Staggered magnetization in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ from ^{139}La NQR and μSR : Effects of Sr doping in the range $0 < x < 0.02$. *Phys. Rev. B* **52**, 7334–7345 (1995).
- Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper-oxide superconductors. *Nature* **375**, 561–563 (1995).
- Dai, P. *et al.* The magnetic excitation spectrum and thermodynamics of high T_c superconductors. *Science* **284**, 1344–1347 (1999).
- Salkola, M., Emery, V. J. & Kivelson, S. A. Implications of charge ordering for single-particle properties of high- T_c superconductors. *Phys. Rev. Lett.* **77**, 155–158 (1996).
- Noda, T., Eisaki, H. & Uchida, S. Evidence for one-dimensional charge transport in $\text{La}_{2-x}\text{Nd}_x\text{Sr}_x\text{CuO}_4$. *Science* **286**, 265 (1999).
- Zaanen, J. & Gunnarsson, O. Charged magnetic domain lines and the magnetism of high- T_c oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
- Emery, V. J., Kivelson, S. A. & Lin, H. Q. Phase separation in the t - J model. *Phys. Rev. Lett.* **64**, 475–478 (1990).
- Schulz, H. J. Incommensurate antiferromagnetism in the two-dimensional Hubbard model. *Phys. Rev. Lett.* **64**, 1445–1448 (1990).
- White, S. R. & Scalapino, D. J. Density matrix renormalization group study of the striped phase in the 2D t - J model. *Phys. Rev. Lett.* **80**, 1272–1275 (1998).
- Emery, V. J., Kivelson, S. A. & Zachar, O. Spin-gap proximity effect mechanism of high-temperature superconductivity. *Phys. Rev. B* **56**, 6120–6147 (1997).
- Zaanen, J., Horsch, M. L. & van Saarloos, W. Charged domain-wall dynamics in doped antiferromagnets and spin fluctuations in cuprate superconductors. *Phys. Rev. B* **53**, 8671–8680 (1996).
- Batlogg, B. *et al.* Charge dynamics in $(\text{La}, \text{Sr})_2\text{CuO}_4$ —from underdoping to overdoping. *J. Low Temp. Phys.* **95**, 23–31 (1994).
- Batlogg, B. & Emery, V. J. Superconductivity—Crossovers in cuprates. *Nature* **382**, 20–21 (1996).
- Homes, C. C., Timusk, T., Liang, R., Bonn, D. A. & Hardy, W. N. Optical conductivity of c -axis oriented $\text{YBa}_2\text{Cu}_3\text{O}_{6.70}$: Evidence for a pseudogap. *Phys. Rev. Lett.* **71**, 1645–1648 (1993).
- Warren, W. W. *et al.* Cu spin dynamics and superconducting precursor effects in planes above T_c in $\text{YBa}_2\text{Cu}_3\text{O}_{6.7}$. *Phys. Rev. Lett.* **62**, 1193–1196 (1989).
- Loeser, A. G. *et al.* Excitation gap in the normal state of underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **273**, 325–329 (1996).
- Hauff, R., Tajima, S., Jang, W. J. & Rykov, A. I. Optical study on phonon anomalies and spin gap in $\text{YBa}_2(\text{Cu}_{1-x}\text{Zn}_x)_3\text{O}_{6.6}$. *Phys. Rev. Lett.* **77**, 4620–4623 (1996).
- Starykh, O. A., Zhitomirsky, M. E., Khomskii, D. I., Singh, R. P. & Ueda, K. Origin of spin gap in CaV_4O_9 : Effects of frustration and lattice distortions. *Phys. Rev. Lett.* **77**, 2558–2561 (1996).
- Bianconi, A. *et al.* Determination of the local lattice distortions in the CuO_2 plane of $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$. *Phys. Rev. Lett.* **76**, 3412–3415 (1996).
- Hellberg, M. & Manousakis, S. Stripes and the t - J Model. *Phys. Rev. Lett.* **83**, 132–135 (1999).
- Kataev, V., Rameev, B., Buchner, B., Hucker, M. & Borowski, R. Slow anti-ferromagnetic dynamics in the low-temperature tetragonal phase of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ as revealed by ESR of Gd spin probes. *Phys. Rev. B* **55**, R3394–R3397 (1997).
- Gemmell, D. S. Channelling and related effects in the motion of charged particles through crystals. *Rev. Mod. Phys.* **46**, 129–227 (1974).
- Lindhard, J. Influence of crystal lattice on motion of energetic charged particles. *Mat. Fys. Medd. Dan. Vid. Selsk.* **34**, 1–64 (1965).
- Barrett, J. H. Monte Carlo channelling calculations. *Phys. Rev. B* **3**, 1527–1547 (1971).
- Sharma, R. P. *et al.* Evidence for large static and dynamic distortions in high T_c superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ crystals over a wide temperature range. *Phys. Rev. Lett.* **77**, 4624–4627 (1996).
- Mook, H. A. & Dogan, F. Charge fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ high-temperature superconductors. *Nature* **401**, 145–147 (1999).
- Deutscher, G. Coherence and single-particle excitations in the high-temperature superconductors. *Nature* **397**, 410–412 (1999).

Acknowledgements

We thank R. L. Greene for reading of the manuscript and suggestions. This work was supported by the National Science Foundation under the Materials Research Science and

Engineering Center and Office of Naval Research at Maryland, by the state of Texas through the Texas Center for Superconductivity at the University of Houston, and the Department of Energy at Argonne.

Correspondence and requests for materials should be addressed to T. V. (e-mail: venky@squid.umd.edu).

Complete photonic bandgaps in 12-fold symmetric quasicrystals

M. E. Zoorob*, M. D. B. Charlton*, G. J. Parker*, J. J. Baumberg*† & M. C. Netti†

* Department of Electronics and Computer Science, † Department of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

Photonic crystals are attracting current interest for a variety of reasons, such as their ability to inhibit the spontaneous emission of light^{1,2}. This and related properties arise from the formation of photonic bandgaps, whereby multiple scattering of photons by lattices of periodically varying refractive indices acts to prevent the propagation of electromagnetic waves having certain wavelengths. One route to forming photonic crystals is to etch two-dimensional periodic lattices of vertical air holes into dielectric slab waveguides^{3–7}. Such structures can show complete photonic bandgaps^{8–10}, but only for large-diameter air holes in materials of high refractive index (such as gallium arsenide, $n = 3.69$), which unfortunately leads to significantly reduced optical transmission when combined with optical fibres of low refractive index. It has been suggested that quasicrystalline (rather than periodic) lattices can also possess photonic bandgaps^{11–14}. Here we demonstrate this concept experimentally and show that it enables complete photonic bandgaps—non-directional and for any polarization—to be realized with small air holes in silicon nitride ($n = 2.02$), and even glass ($n = 1.45$). These properties make photonic quasicrystals promising for application in a range of optical devices^{14–18}.

On the basis of previous theoretical studies^{3,9,10,19,20}, it is evident that the anisotropy of a photonic bandgap (PBG) is dependent on the symmetry of the photonic crystal lattice. As the order of the symmetry increases, the Brillouin zone becomes more circular, resulting in a complete bandgap. The highest level of symmetry found in periodic lattices is six; however, much higher levels of symmetry can be achieved using the more complex geometries of quasicrystals. These quasicrystal photonic lattices eliminate waveguiding modes coincident with the PBG. We investigated photonic quasicrystals based on a random square-triangle tiling system^{21,22}, which possess 12-fold symmetry. These were modelled using a finite difference time domain (FDTD) method to optimally design silicon nitride waveguide structures producing PBGs in the visible/near-infrared spectral region.

The motivation for using a quasicrystal is to maintain the periodic scattering of light while reducing the orientational order of the system so that it is more isotropic. This is achieved using a random ensemble of squares and equilateral triangles to form the lattice of air pores, which are separated by a distance a . In order to exhibit long-range 12-fold symmetry, a large number of cells are required. We used the Stampfli inflation rule, with dodecahedral parent cells (Fig. 1a, dashed lines) generating the offspring dodecagons (Fig. 1a, solid lines). The resulting arrangement of air rods in the waveguide dielectric layers is generated by placing holes at the vertices as shown in Fig. 1b (plan view) and Fig. 1c (section). The photonic quasicrystals examined are composed of 150-nm-diameter air rods

arranged on a pitch a of 260 nm. The 260-nm-thick waveguide of silicon nitride is capped above and below by silicon dioxide to confine the light in the waveguide (Fig. 1c).

Clues to the scattering properties of the quasicrystal may be found in the reciprocal lattice space (Fig. 1d). The large parent cell generates the dodecahedral 'Brillouin zone' indicated, while the edges of Fig. 1d extend to the basic reciprocal vector $2\pi/a$. A triangular or square lattice of the same nearest-neighbour distance would have a sparse reciprocal lattice with points at the corners; however, the quasicrystal has many additional symmetries that average out the anisotropy, but provide additional resonant scattering conditions.

We model the propagation through 28 rows of the structure using an FDTD method. This directly tracks the scattering of waves from a gaussian wavepacket entering the quasicrystal, and is based on a centred difference spatial step²³ to reduce transport errors from scattering in both forward and backward directions. Figure 2a shows the predicted bandgap limits as a function of the air filling fraction β , for a pitch $a = 260$ nm, and TE-polarized light propagating along the Γ -J direction. We note that even for small values of β , a bandgap is apparent. At $\beta = 10\%$, the ratio of the bandgap to the central frequency ($\Delta\omega/\omega_0$) is 13%, whereas for a triangular lattice with the same filling fraction and geometry $\Delta\omega/\omega_0 = 2.5\%$ in the same direction.

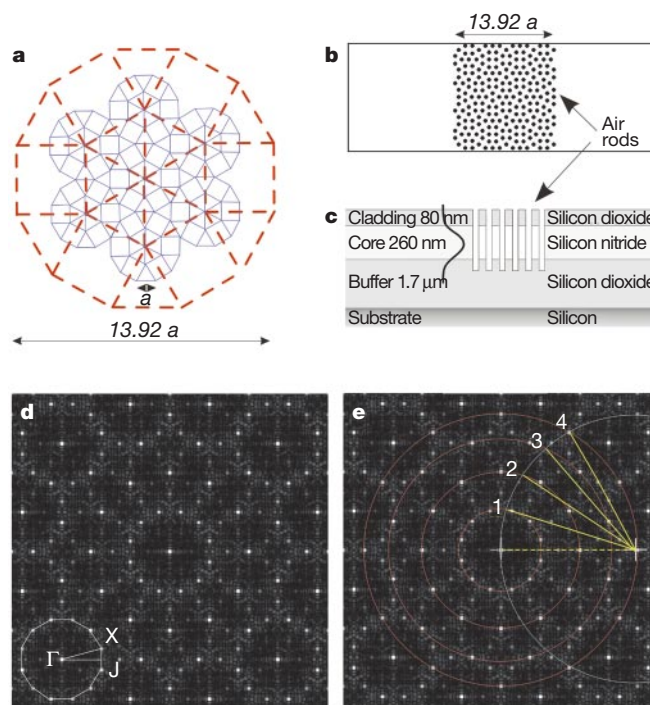


Figure 1 Scheme of the basic quasicrystal elementary unit and its implementation in a planar waveguide. **a**, Quasicrystal pattern based on square-triangular tiling, grown using random Stampfli inflation from parent lattice (dashed red lines) to produce offspring tiling (solid lines). a is the distance between the air rods (pitch) and $13.92a (= a(2 + \sqrt{3}))$ is the total size of the parent lattice. **b**, Plan of air rods placed at quasicrystal vertices and etched through the planar waveguide. **c**, Cross-section of waveguide structure, consisting of 260-nm-thick silicon nitride ($n = 2.04$) sandwiched between two silicon dioxide ($n = 1.46$) layers grown on a thick silicon buffer layer. The depth of holes and calculated waveguide mode in planar regions are also shown. **d**, Representation of the quasicrystal reciprocal lattice, showing 12-fold symmetry. The Brillouin zone (white dodecagon) corresponds to the supercell shown dashed in **a**. **e**, Ewald sphere construction for the 633-nm light used in the experiment. k_{in} (dashed) defines the radius of the sphere (grey) whose intersection with planes of air rods (red circles in reciprocal space) gives the diffraction directions k_{in} (marked 1–4).

Transmission spectra were calculated in all directions using the FDTD method for a photonic quasicrystal with $a = 260$ nm and $\beta = 28\%$. Figure 2b shows the predicted transmission for both TE- and TM-polarized modes along the Γ -J direction. We use the normalized frequencies a/λ for conveniently comparing different structures. The complete PBG persists from $0.158a/\lambda$ to $0.190a/\lambda$ in all directions for the first bandgap. A higher-order bandgap is also seen from $0.408a/\lambda$ to $0.418a/\lambda$. The normalized width of the primary bandgap is $\Delta\omega/\omega_0 = 33.0\%$ for TE modes, and $\Delta\omega/\omega_0 = 20.5\%$ for TM modes. In contrast to regular lattices, the TM PBG lies exactly in the middle of the TE PBG²⁴. This gives rise to a 65.5% overlap in the complete PBG for the two different polarization states, compared to the total absence of a complete PBG for triangular and hexagonal lattice structures with the same air filling fraction³. In the long-wavelength limit ($a/\lambda < 0.10$), the transmission reaches unity. At these wavelengths the fine structure is not resolved by the propagating waves, and the quasicrystal is seen as homogeneous material of reduced refractive index.

The quasicrystal bandgaps occur at lower frequencies than in the

triangular and hexagonal crystals. The normalized midgap frequencies for TE and TM are $0.170a/\lambda$ and $0.176a/\lambda$, respectively, compared to $\sim 0.375a/\lambda$ for regular lattices²⁴. In contrast to regular lattice structures, nearest-neighbour lattice vectors no longer define the 'Brillouin zone'. As a consequence of quasicrystal symmetry, the 'Brillouin zone' is instead defined by lattice vectors joining air rods several periods away, as seen by the 'Brillouin zone' in Fig. 1d. Further simulations show that the photonic bandgap remains open for very low index materials such as glass ($n = 1.45$) with an air filling fraction β of only 30%. The calculated transmission shown in Fig. 2c confirms that it should be possible in practice to integrate photonic quasicrystal devices with optical fibre systems.

Quasicrystal waveguides have been fabricated with a range of parameters, and demonstrated good reproducibility and stability. Figure 3a shows an scanning electron micrograph of a typical device from which the pitch of 260 nm and a fill fraction of 28% are measured. To confirm the enhanced diffraction predicted from the quasicrystal reciprocal lattice, the far-field in-plane diffraction is shown in Fig. 3b. TE-polarized light at a wavelength of 633 nm enters from a cleaved waveguide face on the lower edge, and propagates up to the horizontal line of the photonic quasicrystal. Many diffracted beams are observed, which can be compared to simple theory using the Ewald sphere construction in Fig. 1e. The Bragg scattering condition $\mathbf{k}_{\text{out}} = \mathbf{k}_{\text{in}} + \mathbf{G}$, where $\mathbf{G}$ is a reciprocal lattice vector, can be represented by a circle centred at $\mathbf{k}_{\text{in}} = 2\pi/\lambda_{\text{in}}$, which indicates photon scattering where it intersects the reciprocal lattice points. Note that it is necessary to take into account the

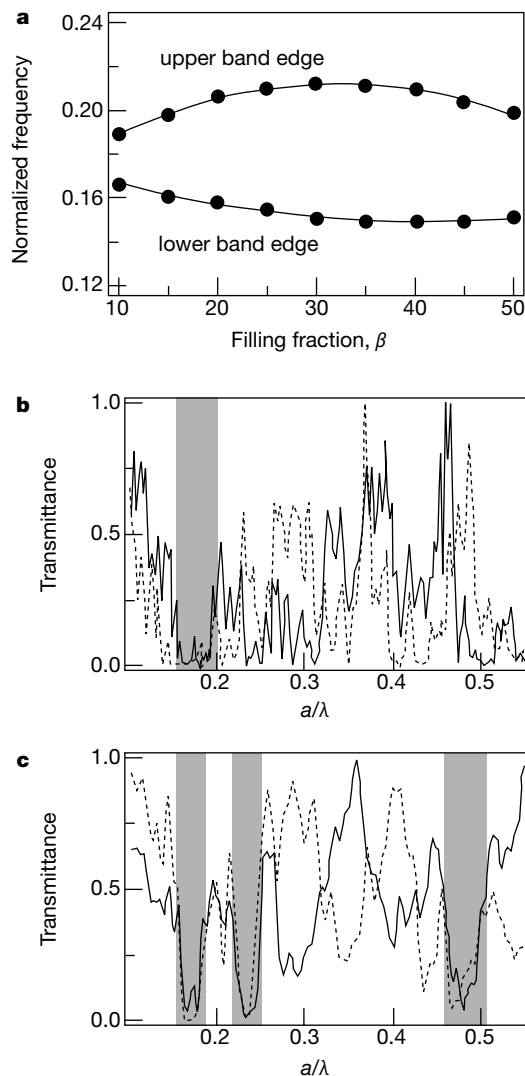


Figure 2 Theoretical calculation. **a**, Photonic bandgap extrema as a function of filling fraction, β , for $a = 260$ nm with light along Γ -J. The gaps vary by less than 2% in different directions. **b**, Transmission spectra for TE (dashed line) and TM (solid line) polarizations as a function of normalized frequencies a/λ ($a = 260$ nm), calculated for a quasicrystal in SiN waveguide having air-filling fraction $\beta = 28\%$, and **c**, for the same lattice in glass with $\beta = 30\%$. The shaded areas mark the main gaps and their superposition for both TE and TM modes.

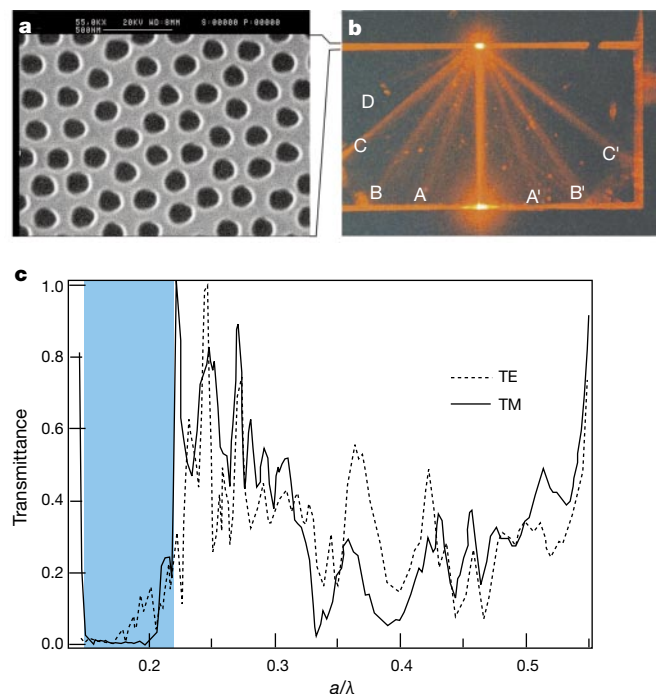


Figure 3 Experimental results. **a**, SEM image of a 12-fold symmetric photonic quasicrystal structure. **b**, Photograph demonstrating the multiple beam diffracted reflections at the entrance face of the photonic quasicrystal. The illuminating laser $\lambda = 633$ nm, compared to the photonic quasicrystal pitch of $a = 260$ nm and a filling fraction of $\beta = 0.28$. The angles of the diffracted beams A'-C' and A-D correspond to the 1-4 diffraction directions calculated by means of the Ewald sphere (Fig. 1e). The angle values are reported in Table 1. **c**, Experimental transmission in a 12-fold photonic quasicrystal with $a = 260$ nm, $\beta = 28\%$ along Γ -J. The laser system used in these measurements allowed us to cover a spectrum from 450 to 1,800 nm. The curves are corrected to account for short-wavelength scattering from the waveguide using the normalization functions $\exp[-(850\text{nm}/\lambda)^3]$ for TE and $\exp[-(900\text{nm}/\lambda)^2]$ for TM. The blue area highlights the gaps for both TE (dashed line) and TM (solid line) polarization.

effective refractive index of the etched area, which affects the size of the Ewald sphere and hence the $\mathbf{k}_{\text{out}}$ direction. This orientation-dependent effective index is an approximation, which is only valid in the near-isotropic quasicrystals. The comparison between observed and predicted angles of diffraction is reported in Table 1, and it shows reasonable agreement. Thus, the scattered wavevectors clearly demonstrate the 12-fold symmetry produced by the quasicrystal nanostructures and confirms their successful fabrication.

Transmission measurements were performed by spectrally resolving the light propagating through the lattice in different directions. The spectra reported in Fig. 3c show (we believe for the first time) experimental evidence for complete bandgaps in quasicrystals for both TE and TM polarizations. The experiments used a white light continuum produced by focusing pulses (1 μJ , 100 fs) from a regenerative amplifier tuned to 850 nm into 1 mm of sapphire. The availability of such high-brightness ultra-broadband laser sources facilitates high-accuracy transmittance measurements from 450 to 1,800 nm through the waveguide. The use of achromatic optics and a carefully designed optical fibre spatial filter provided excellent collimation and pointing properties for coupling into the planar waveguide. This special optical fibre also provided extra continuum generation in the range 0.8 to 1.8 μm for measurements in the near-infrared. Because the air holes also cause scattering losses, the spectra in Fig. 3c have been normalized to account approximately for the decreased throughput at shorter wavelengths.

The main features of the transmission spectra in Fig. 3c are the presence of sharp gaps amid strongly modulated structure, very similar to that predicted by the simulations. This differs dramati-

Table 1 Experimental and theoretical diffraction angles

Experimental beam*	Measured angle	Theoretical beam†	Theoretical angle
A, A'	20.9, 18.8	1	18.0
B, B'	36.3, 33.0	2	34.6
C, C'	51.3, 50.1	3	51.3
D	62.0	4	62.5

Shown are experimental and theoretical angles of diffracted reflections from a photonic quasicrystal with $a = 260$ nm and $\lambda = 633$ nm.

* Fig. 3b.

† Fig. 1e.

cally from regular lattice structures, which exhibit less spectral structure in keeping with their sparser reciprocal lattices²⁴. The primary bandgap for both TE and TM modes are exactly centred at $0.17a/\lambda$, showing an extinction ratio as good as 10^{-3} , in excellent agreement with the theory. In addition, the experimental transmittances show higher-order bandgaps for both TE and TM polarized gaps centred at $0.352a/\lambda$ and $0.413a/\lambda$ with an extinction ratio of about 90%.

The experimental investigation of the angular dependence of the main PBG at $0.17a/\lambda$ is shown in Fig. 4a. The lower and upper transmission spectra were obtained at normal incidence and 6° around the Γ -J direction, respectively. The limit of the sensitivity of the infrared detector is at about $0.14a/\lambda$. The theoretical simulations along the same directions are given in Fig. 4b, showing a very good agreement with the experimental spectra.

Reproducibility and fabrication tolerances were also investigated by sampling the transmission characteristics along the photonic quasicrystal device at several different positions (spaced by ~ 1 mm) as shown in Fig. 4c. The fine spectral details of all the different transmission spectra are maintained, indicating excellent fabrication quality as confirmed by Fig. 3a.

All experimental spectra show the same fine structure as that calculated in Fig. 2b, and in particular the predicted angular and polarization dependence. Although the primary PBG width is reduced by a factor of 1.05 compared to theory, the central frequency is identical. This small mismatch is not surprising if one takes into account that the FDTD calculation is a two-dimensional model that neglects the real three-dimensional character of waveguiding in a planar structure. This discrepancy between the theoretical and experimental results clearly demonstrates that although the refractive index periodicity has a two-dimensional translational symmetry, a full three-dimensional treatment needs to be pursued for these structures. This is the case even for regular photonic crystals, where recent experimental and theoretical results show similar discrepancies in transmission spectra^{5,24–26}. Full three-dimensional theoretical modelling of these waveguiding quasicrystal devices is a non-trivial task, given the lack of long range periodicity and the added complexity of satisfying waveguiding conditions. Promising results are predicted on the basis of previous initial investigations on the three-dimensional modelling of periodic structures²⁷. This work is currently in progress, and as shown here, an appropriate assembly of air cylinders in a low-refractive-index-dielectric waveguide is a promising technology for controlling the flow of light. □

Received 16 December 1999; accepted 28 February 2000.

- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals* (Princeton Univ. Press, New York, 1995).
- Charlton, M. D. B. & Parker, G. J. Guided mode analysis, and fabrication of a 2-dimensional visible photonic band structure confined within a planar semiconductor waveguide. *Mater. Sci. Eng. B* **49**, 155–165 (1997).
- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
- Gadot, F. *et al.* Experimental demonstration of complete photonic bandgap in graphite structure. *Appl. Phys. Lett.* **71**, 1780–1782 (1997).
- Krauss, T. F., De La Rue, R. M. & Brand, S. Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
- Atkin, D. N., Russell, P. S. J., Birks, T. A. & Roberts, P. J. Photonic band structure of guided Bloch

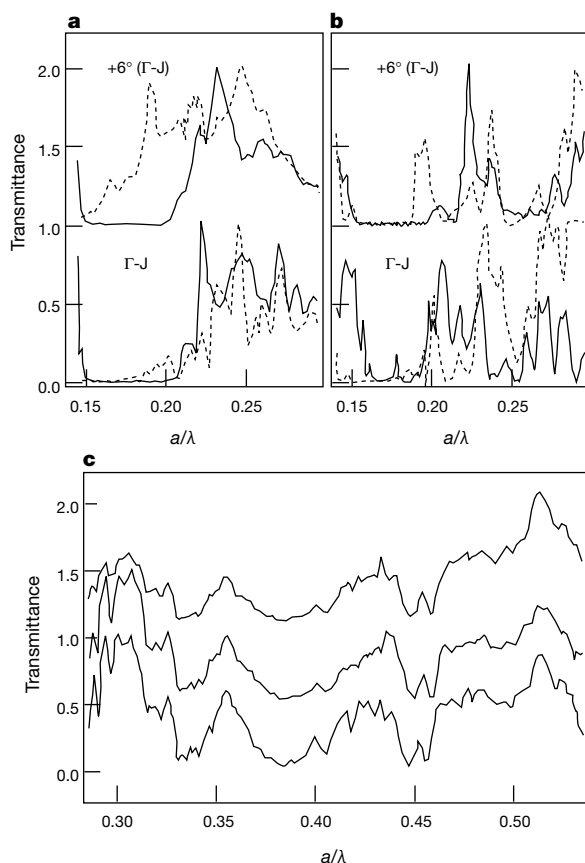


Figure 4 Angular dependence of the primary PBG and reproducibility of transmittance. **a**, Experimental transmission along the Γ -J direction (lower pair of curves) and 6° between Γ -J and Γ -X directions (upper pair of curves) for both TE (dashed lines) and TM (solid lines) modes. **b**, Theoretical transmission for the same angles and frequency range as in **a**. **c**, Scanning of transmission at different points (~ 1 -mm spacing) along the device.

- modes in high index films fully etched through with periodic microstructure. *J. Mod. Opt.* **43**, 1035–1053 (1996).
8. Pilihal, M. & Maradudin, A. A. Photonic band-structure of 2-dimensional systems—the triangular lattice. *Phys. Rev. B* **44**, 8565–8571 (1991).
 9. Cassagne, D., Jouanin, C. & Bertho, D. Hexagonal photonic-band-gap structures. *Phys. Rev. B* **53**, 7134–7142 (1996).
 10. Barra, A., Cassagne, D. & Jouanin, C. Existence of two-dimensional absolute photonic band gaps in the visible. *Appl. Phys. Lett.* **72**, 627–629 (1998).
 11. Chan, Y. S., Chan, C. T. & Liu, Z. Y. Photonic band gaps in two dimensional photonic quasicrystals. *Phys. Rev. Lett.* **80**, 956–959 (1998).
 12. Cheng, S. S. M., Li, L., Chan, C. T. & Zhang, Z. Q. Defect and transmission properties of two-dimensional quasiperiodic photonic band-gap systems. *Phys. Rev. B* **59**, 4091–4099 (1999).
 13. Krauss, T. F. & De la Rue, R. M. Photonic crystals in the optical regime—past, present and future. *Prog. Quant. Electron.* **23**, 51–96 (1999).
 14. Charlton, M. D. B., Parker, G. J. & Zoorob, M. E. Recent developments in the design and fabrication of visible photonic band gap waveguide devices. *J. Mater. Sci.* **10** (Materials in Electronics), 429–440 (1999).
 15. Foresi, J. S. *et al.* Photonic-bandgap microcavities in optical waveguides. *Nature* **390**, 143–145 (1997).
 16. Temelkuran, B. & Ozbay, E. Experimental demonstration of photonic crystal based waveguides. *Appl. Phys. Lett.* **74**, 486–488 (1999).
 17. Kosada, H. *et al.* Superprism phenomena in photonic crystals. *Phys. Rev. B* **58**, R10096–R10099 (1998).
 18. Ohetera, Y., Sato, T., Kawashima, T., Tamamura, T. & Kawakami, S. Photonic crystal polarisation splitters. *Electron. Lett.* **35**, 1271–1272 (1999).
 19. McGurn, A. R. & Maradudin, A. A. Weak transverse localisation of light scattered incoherently from a one-dimensional random metal-surface. *J. Opt. Soc. Am B* **10**, 539–545 (1993).
 20. Anderson, C. M. & Giapis, K. P. Symmetry reduction in group 4mm photonic crystals. *Phys. Rev. B* **56**, 7313–7320 (1997).
 21. Zoorob, M. E., Charlton, M. D. B. & Parker, G. J. *Proc. Inst. Phys. PREP* **99** 161–164 (1999).
 22. Oxborrow, M., Henley, C. L. Random square-triangle tilings: A model for twelfold-symmetric quasicrystals. *Phys. Rev. B* **48**, 6966–6998 (1993).
 23. Yee, K. S. Numerical solutions of initial boundary value problems involving Maxwell's equation in isotropic media. *IEEE Trans. Antennas Propagat.* **AP-14**, 302–307 (1966).
 24. Netti, M. C., Charlton, M. D. B., Parker, G. J. & Baumberg, J. J. Visible photonic bandgap engineering in silicon nitride waveguides. *Appl. Phys. Lett.* (in the press).
 25. Feng, X. -P. & Arakawa, Y. Off-pane angle dependence of photonic band gap in a two-dimensional photonic crystal. *IEEE J. Quantum. Electron.* **32**, 535–542 (1996).
 26. Labilloy, D. *et al.* Quantitative measurement of transmission, reflection, and diffraction of two-dimensional photonic band gap structures at near infrared wavelengths. *Phys. Rev. Lett.* **79**, 4147–4150 (1997).
 27. Ho, K. M., Chan, C. T., Soukoulis, C. M., Biswas, R. & Sigalis, M. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Commun.* **89**, 413–416 (1994).

Acknowledgements

We thank the holey-fibre group of the Southampton Optoelectronics Research Centre for supplying samples. This work was supported by the EPSRC, the HEFCE and the University of Southampton.

Correspondence and requests for materials should be addressed to G.J.P. (e-mail: gjp@ecs.soton.ac.uk).

Manipulation of atoms across a surface at room temperature

T. W. Fishlock*, A. Oral*†, R. G. Egddell‡ & J. B. Pethica*

* Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

‡ Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK

Since the realization that the tips of scanning probe microscopes can interact with atoms at surfaces, there has been much interest in the possibility of building or modifying nanostructures or molecules directly from single atoms¹. Individual large molecules can be positioned on surfaces^{2–4}, and atoms can be transferred controllably between the sample and probe tip^{5,6}. The most complex structures^{7–11} are produced at cryogenic temperatures by sliding atoms across a surface to chosen sites. But there are problems in manipulating atoms laterally at higher tempera-

tures—atoms that are sufficiently well bound to a surface to be stable at higher temperatures require a stronger tip interaction to be moved. This situation differs significantly from the idealized weakly interacting tips^{12,13} of scanning tunnelling or atomic force microscopes. Here we demonstrate that precise positioning of atoms on a copper surface is possible at room temperature. The triggering mechanism for the atomic motion unexpectedly depends on the tunnelling current density, rather than the electric field or proximity of tip and surface.

For Br adsorbed on Cu(001), we find that whereas normal stable and atom-resolved scanning tunnelling microscope (STM) images can be obtained at lower tunnel currents, the adsorbed atoms start moving if the images are taken with a tunnel current above a few nanoamps. The tunnel current provides a remarkably well defined motion-control parameter. The STM images reported here were taken at room temperature with an Omicron ultrahigh vacuum (UHV) STM. The line scans are horizontal, and start at the bottom of the image. The Cu(001) was cleaned by standard UHV procedures and the bromine was then deposited at low coverages using an *in situ* electrochemical cell doser¹⁴.

Figure 1 shows an image where the tunnel current is changed from 3 to 1 nA halfway up the scan. In the first, lower part of the image (3 nA), most of the Br adsorbates appear as a series of tracks, perpendicular to the line scan. In the upper (1 nA) part of the image, the Br atoms are in fixed, well defined locations on the surface. The tracks in the lower part of the image are in fact single atoms being driven in front of the tip, and hopping from one Cu surface site to another. Figure 2 is the same surface region just afterwards, but with a current of 1 nA, which is below the threshold to induce motion. The atoms marked ABCD are seen to be still in the same place as they were left in Fig. 1, but the area below them has been 'swept' clear of most Br adsorbates. For 95% of our images, the underlying metal lattice has a corrugation too small to observe, as is expected for low-index metal surfaces¹⁵; but occasionally a tip structure occurs which enables the metal atoms to be seen¹⁵. We find that the tip structure makes no significant difference to the general character of the Br motion we observe. Therefore, in the remaining figures we show these rather rarer types of image where the metal lattice can be seen, to make the local nature of the atom motion completely clear.

The image shown in Fig. 3, taken at 2 nA, clearly shows the underlying Cu(001) lattice registry. Most atoms are hopping

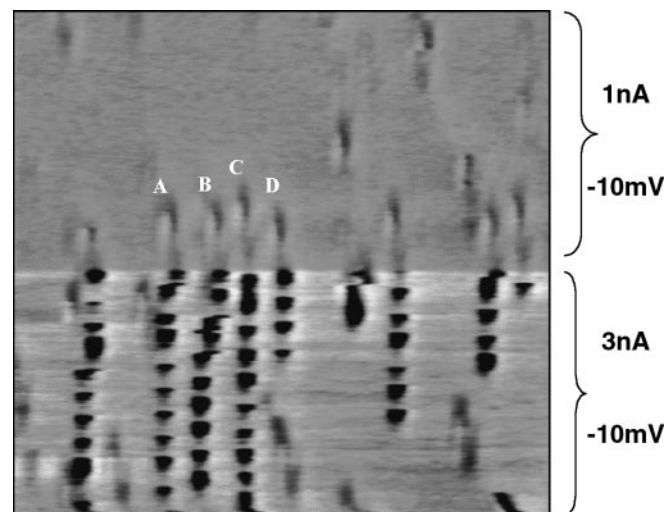


Figure 1 STM image, $60 \text{ \AA} \times 60 \text{ \AA}$, of Br on Cu(100). The tunnel current is reduced halfway up the image. The tracks in the 3-nA (lower half) area are single atoms hopping from site to site. When the current is reduced the bromine atoms stop moving, for example those marked A,B,C and D.

† Present address: Department of Physics, Bilkent University, 06533 Ankara, Turkey.

between nearest-neighbour sites in the $\{110\}$ direction most nearly normal to the line scan direction. The line scan in Fig. 1 is along $[110]$, producing essentially straight tracks, whereas in Fig. 3 it is slightly misoriented from $[110]$, giving zigzag tracks—most individual hops are along $[1\bar{1}0]$ but some are along $[110]$. Aligning image line scans along $\{100\}$ produces fairly even zigzag tracks at 45° to the scan line. The form of the hopping motion of a single atom is shown in more detail in a line section of the image along one of the hopping paths (Fig. 4). The bromine is imaged as a depression in the surface, so as the tip nears the atom its track lowers towards the surface. At a critical separation the Br jumps away from the tip, so the tip then jumps back out to the normal height above the Cu surface. The process is similar that described by Meyer *et al.*¹⁰ for carbon monoxide on Cu at 30 K.

We conclude from all the above that the atoms are pushed (repelled) by the tip along the easiest available $\{110\}$ direction. As might be expected, this is the direction with the lowest diffusion barrier to motion. The motion resembles a 'cut' or 'sliced' shot in billiards. The atom moves at nearly right angles to the direction of tip scan, because the tip's closest-approach distance to the Br atom increments in the direction normal to the raster line scan.

It might be thought that all that is now required to move an atom to a specified location is to move the tip towards the atom from the opposite direction with a tunnel current above the threshold. This is much more difficult at room temperature than at cryogenic temperatures, because minute drifts of tip position between the time of the prior imaging (below threshold, to see what is to be manipulated) and the manipulation stroke itself, cause the tip to approach off-centre to the atom—and hence to risk either an undesired sideways jump or even a complete miss.

We have solved this problem by modifying the motion-inducing stroke to include a small rapid 'dither' perpendicular to the stroke direction. The stroke thus locally replicates the effect of the image line scan on a single atom. Figure 5 shows the 'before' and 'after' images of such a stroke on the marked atom. It has been deliberately displaced by five lattice spacings, and this can be done in a controlled manner over any chosen, programmed distance. Clearly, the size of the lateral dither affects the proximity to which groups of atoms can be laterally positioned—at present, this is within a couple of nearest-neighbour spacings. A thermally highly stable STM, with a smaller and more carefully shaped lateral dither

pattern, should permit very close atom placement at room temperature and above.

For a given tip, the threshold current required to induce atom motion was remarkably stable and reproducible. Even for different tips, and hence different appearances of the metal and Br atoms, the value always lay between 1 and 6 nA. We found that the tips that had the highest threshold currents also gave the smallest adsorbate image corrugations below the threshold. Small image corrugation suggests blunter tips, hence a larger area over which the tunnel current is spread and thus a lower current density at a single atom site. This suggests that the threshold at an atom might be governed by current density at that atom rather than just total tip-surface current.

Varying the tunnel parameters produced some interesting effects. (1) The polarity of voltage between tip and surface had no discernible effect on the threshold current for motion: the polarity could even be reversed during motion, with no apparent effect. This implies that fixed dipoles at the surface have no significant role. (2) Increasing the imaging voltage also did not detectably affect the threshold current. Images were taken at 1 nA and tip biases varying between 5 mV and 400 mV, and at none of these biases was any atom motion induced. The forces on induced atom dipoles depend on field squared, so there should be a very substantial change in tip-field-induced polarization forces over this voltage range, but nothing was in fact seen. (3) In contrast, in the same experiments, if the current was increased to 3 nA, the atoms always moved, irrespective of tip voltage. Taken together, these results strongly imply that the electric field between tip and surface is not directly responsible for the observed induction of motion.

More surprisingly, halving the voltage at constant current during an image failed to induce motion, whereas doubling the current at constant voltage during the same image did induce motion. Because in both cases the tip-surface separation should be reduced by a roughly similar amount, this implies that the actual distance between tip and surface itself is not solely responsible for the onset of atom movement. That is, although there is surely an overall distance-dependent force^{12,13} with some repulsion on the Br atom, such short-range forces seem not to be the dominant factor in triggering individual site-to-site lateral motion. This leaves us with the actual tunnel current density as the main mechanism for

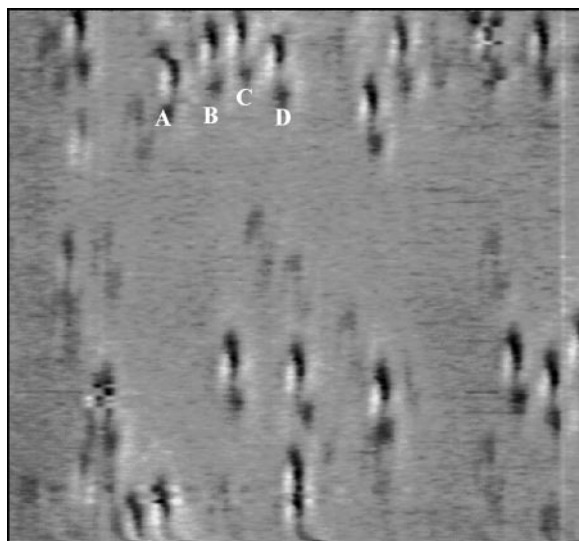


Figure 2 Image taken immediately after Fig. 1, showing the same atoms marked A, B, C and D. The area below them, swept almost clear of Br atoms, is the 3-nA area of Fig. 1.

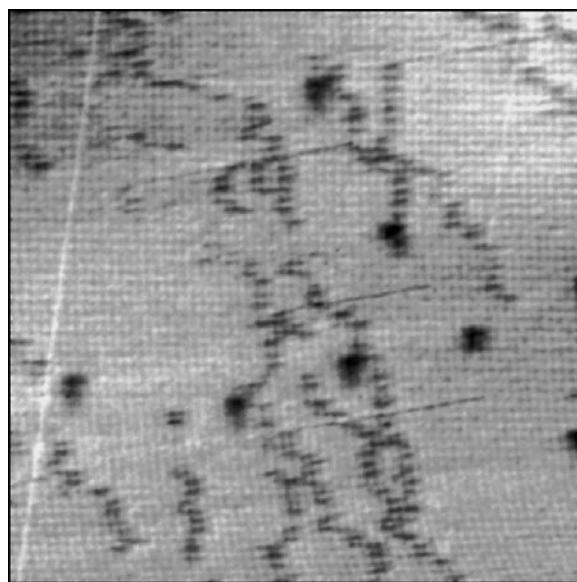


Figure 3 STM image, $120 \text{ \AA} \times 120 \text{ \AA}$, showing most Br atoms moving across the clearly visible Cu lattice. (The image was taken at 2 nA, 10 mV.) Each individual hop is in a $\{110\}$ direction.

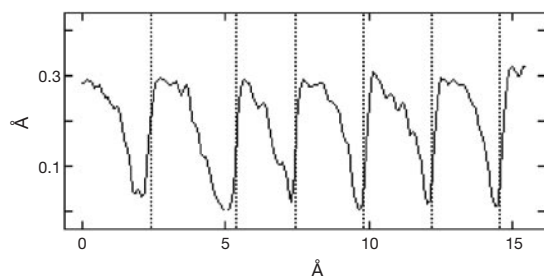


Figure 4 Vertical section through an image along the track of a single moving atom. The individual jumps between adjacent adsorption sites are shown. The atom motion is left to right.

inducing motion. We stress again that none of the above observations or conclusions appear to be affected by the tip type or structure.

For two images taken of the same area at low current, with a delay of a few minutes between them, we observe that some 5–10% of the Br atoms have moved, typically by one site step. Assuming an Arrhenius prefactor of about 10^{13} Hz, the lateral barrier to motion can therefore be crudely estimated to be a little under 1 eV. If we raise the sample temperature by 80 °C or so for even a short time, the Br atoms all diffuse away to steps and other sinks, which also suggests that the lateral diffusion barrier is around 1 eV. Now therefore if an energy of around 10 meV, additional to room temperature energy, can be delivered to an atom from an increased tunnelling current, the hopping rate of this atom will increase to a second or so per hop. With the addition of a background tip–atom repulsion, this is sufficiently rapid to account for the manipulation we actually observe.

Following previous workers^{16,17}, we therefore suggest that the tunnel current is inelastically scattered, giving vibrational excitation (heating) of the Br. The excess temperature reached by the Br atom

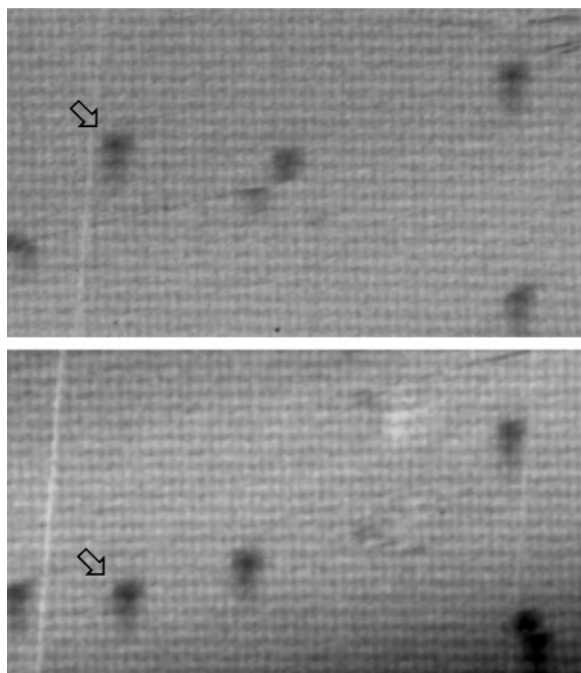


Figure 5 Controlled positioning of a single atom at room temperature. The lower image is taken before, and the upper image after, a manipulation stroke on the arrowed atom. Note that the next Br atom to the right moved during the second image. Images were taken at 1 nA and 10 mV; the manipulation between images was at 3 nA.

would then depend on the balance between the vibrational relaxation rate and the energy input rate^{16–18}. The latter is proportional to the current, giving the behaviour we observe. An increase in current by a factor of two causes the number of adsorbed atoms moving during a single image to go from a very small fraction to more than 80%. The inelastic scattering cross-section would not vary greatly over our range of a few hundred millivolts, so that the motion onset should be largely independent of voltage and in particular of polarity, as observed. Note that when we vary the voltage in the experiments, the current remains constant. We did not use voltages below 10 meV in the motion experiments, so there is always enough energy available to excite the Br motion. Multiple excitation before relaxation is also possible^{17,19}, which might allow use of even lower voltages, but we do not have direct evidence to indicate this at present. We point out that all surface atoms become very unstable in general STM when the currents reach values around 100 nA.

A variety of mechanisms of atom manipulation have now been reported^{11,19,20}. The type of mechanism acting in a particular case depends on the atom species being manoeuvred, and more than one threshold process might be present in more complex chemical and molecular systems. Atom positioning is clearly possible at elevated temperatures, and we speculate that some chemical specificity of manipulation mechanism might become possible, leading eventually to atom selectivity when manipulating more complex molecules or structures. □

Received 1 September 1999; accepted 9 February 2000.

1. Avouris, Ph. Manipulation of matter at the atomic and molecular-levels. *Acc. Chem. Res.* **28**, 95–102 (1995).
2. Jung, T. A., Schittler, R. R., Gimzewski, J. K., Tang, H. & Joachim, C. Controlled room-temperature positioning of individual molecules: Molecular flexure and motion. *Science* **271**, 181–184 (1996).
3. Cuberes, M. T., Schittler, R. R. & Gimzewski, J. K. Room-temperature repositioning of individual C_{60} molecules at Cu steps: Operation of a molecular counting device. *Appl. Phys. Lett.* **69**, 3016–3018 (1996).
4. Beton, P. H., Dunn, A. W. & Moriarty, P. Manipulation of C_{60} molecules on a Si surface. *Appl. Phys. Lett.* **67**, 1075–1077 (1995).
5. Lyo, I.-W. & Avouris, Ph. Field-induced nanometer-scale to atomic-scale manipulation of silicon surfaces with the STM. *Science* **253**, 173–176 (1991).
6. Dujardin, G. *et al.* Vertical manipulation of individual atoms by a direct STM tip-surface contact on Ge(111). *Phys. Rev. Lett.* **80**, 3085–3088 (1998).
7. Crommie, M. F., Lutz, C. P., Eigler, D. M. & Heller, E. J. Quantum corrals. *Physica D* **83**, 98–108 (1995).
8. Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunneling microscope. *Nature* **344**, 524–526 (1990).
9. Whitman, L. J., Strosio, J. A., Dragoset, R. A. & Cellota, R. Manipulation of adsorbed atoms and creation of new structures on room-temperature surfaces with a scanning tunneling microscope. *Science* **251**, 1206–1210 (1991).
10. Bartels, L., Meyer, G. & Rieder, K.-H. Basic steps of lateral manipulation of single atoms and diatomic clusters with a scanning tunneling microscope tip. *Phys. Rev. Lett.* **79**, 697–700 (1997).
11. Meyer, G., Zophel, S. & Rieder, K.-H. Manipulation of atoms and molecules with a low temperature scanning tunneling microscope. *Appl. Phys. A* **63**, 557–564 (1996).
12. Jarvis, S. P., Yamada, H., Yamamoto, S.-I., Tokumoto, H. & Pethica, J. B. Direct mechanical measurement of interatomic potentials. *Nature* **384**, 247–249 (1996).
13. Cross, G. *et al.* Adhesion interaction between atomically defined tip and sample. *Phys. Rev. Lett.* **80**, 4685–4688 (1998).
14. Spencer, N. D., Goddard, P. J., Davies, P. W., Kitson, M. & Lambert, R. M. A simple, controllable source for dosing molecular halogens in u.h.v. *J. Vac. Sci. Technol. A* **1**, 1554–1555 (1983).
15. Wintterlin, J. *et al.* Atomic resolution imaging of close-packed metal surfaces by scanning tunneling microscopy. *Phys. Rev. Lett.* **62**, 59–62 (1989).
16. Ralls, K. S., Ralph, D. C. & Buhrman, R. A. Individual-defect electromigration in metal nanobridges. *Phys. Rev. B* **40**, 11561–11570 (1989).
17. Walkup, R. E., Newns, D. M. & Avouris, Ph. Role of multiple inelastic transitions in atom-transfer with the scanning tunneling microscope. *Phys. Rev. B* **48**, 1858–1861 (1993).
18. Todorov, T. N. Local heating in ballistic atomic-scale contracts. *Phil. Mag.* **77**, 965–973 (1998).
19. Shen, T. C. *et al.* Atomic-scale desorption through electronic and vibrational-excitation mechanisms. *Science* **268**, 1590–1592 (1995).
20. Hansen, O., Ravnkilde, J. T., Quaade, U., Stokbro, K. & Grey, F. Field-induced deformation as a mechanism for scanning tunneling microscopy based nanofabrication. *Phys. Rev. Lett.* **81**, 5572–5575 (1998).

Acknowledgements

We thank M. Welland, T. Todorov and D. Eigler for comments. The apparatus used in this work was originally funded by EPSRC.

Correspondence and requests for materials should be addressed to J.B.P. (e-mail: john.pethica@materials.ox.ac).

Self-assembly of nanoparticles into structured spherical and network aggregates

Andrew K. Boal*, Faysal Ilhan*, Jason E. DeRouchey†, Thomas Thurn-Albrecht†, Thomas P. Russell† & Vincent M. Rotello*

*Department of Chemistry and †Department of Polymer Science and Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA

Multi-scale ordering of materials is central for the application of molecular systems^{1–3} in macroscopic devices^{4,5}. Self-assembly based on selective control of non-covalent interactions^{6–8} provides a powerful tool for the creation of structured systems at a molecular level, and application of this methodology to macromolecular systems provides a means for extending such structures to macroscopic length scale^{9–11}. Monolayer-functionalized nanoparticles can be made with a wide variety of metallic and non-metallic cores, providing a versatile building block for such approaches. Here we present a polymer-mediated ‘bricks and mortar’ strategy for the ordering of nanoparticles into structured assemblies. This methodology allows monolayer-protected gold particles to self-assemble into structured aggregates while thermally controlling their size and morphology. Using 2-nm gold particles as building blocks, we show that spherical aggregates of size 97 ± 17 nm can be produced at 23 °C, and that 0.5–1 µm spherical assemblies with $(5–40) \times 10^5$ individual subunits form at –20 °C. Intriguingly, extended networks of ~50-nm subunits are formed at 10 °C, illustrating the potential of our approach for the formation of diverse structural motifs such as wires and rods. These findings demonstrate that the assembly process provides control over the resulting aggregates, while the modularity of the ‘bricks and mortar’ approach allows combinatorial control over the constituents, providing a versatile route to new materials systems.

Colloids functionalized with self-assembled monolayers (SAMs) are nanoscopic entities that can provide interchangeable ‘building blocks’ for the fabrication of microscale constructs^{12–16}. But conventional synthetic procedures greatly limit the scale of discrete systems that can be produced. Self-assembly provides a powerful alternative to these procedures, potentially allowing the biomimetic

construction of large, complex systems. Mirkin and co-workers^{17,18} and Alivasatos *et al.*¹⁹ have demonstrated the formation of aggregated metal clusters, using DNA as the recognition element. These aggregates, however, were either small in size or irregular in shape, possibly limiting their use in multi-scale devices.

To provide a general means for the controlled self-assembly of nanoparticles, we have developed a ‘bricks and mortar’ approach. In this strategy, colloidal gold particles functionalized with recognition elements serve as the bricks²⁰, while polymers bearing complementary functionality serve as mortar²¹, holding together the colloidal particles. Using this strategy, the conformational flexibility of the polymer compensates for irregularities in the size and shape of the aggregate structure, allowing the efficient propagation of order during the self-assembly process.

Complementarity between colloid and polymer was achieved using the diaminotriazine-thymine three-point hydrogen bonding interaction (Fig. 1a)^{22,23}. For the polymer component, diaminotriazine-functionalized polystyrene was employed (compound 1 in Fig. 1b)²¹. The required thymine-functionalized colloids were synthesized starting with ~2-nm gold particles covered with an octanethiol SAM²⁴. Thiol place-exchange²⁵ with a thymine-functionalized alkanethiol then provided the derivatized colloid ‘Thy-Au’ (Fig. 2). To provide a control system that cannot participate in hydrogen bonding, the analogous ‘MeThy-Au’ (Fig. 2) was prepared in a similar fashion.

Addition of polymer 1 to concentrated solutions of Thy-Au in non-competitive solvents such as dichloromethane and chloroform resulted in rapid formation of a black solid. In contrast, no precipitation was observed upon addition of 1 to the control colloid MeThy-Au. The lack of aggregation observed in the control system demonstrates that the precipitation observed upon the addition of 1 to Thy-Au was the result of specific hydrogen-bond interactions between the colloid thymines and the polymer triazines.

To provide control of the aggregation process, polymer 1 and Thy-Au were mixed in dilute dichloromethane solution, with precipitation proceeding over a 96-h period. The resulting solid was insoluble in non-polar solvents, but >80% could be dissolved in polar media such as methanol and tetrahydrofuran (THF). Small-angle X-ray scattering (SAXS) was used to characterize the

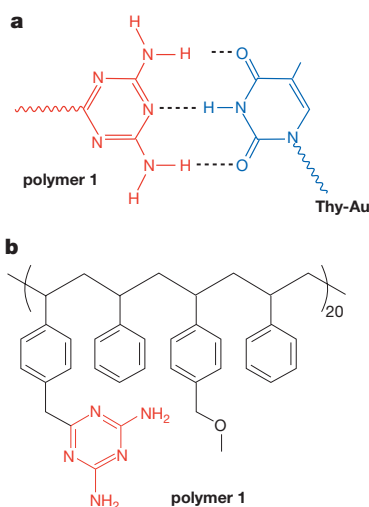


Figure 1 Recognition motif and polymer ‘mortar’. **a**, Diaminotriazine-thymine recognition; **b**, triazine-functionalized random copolymer 1.

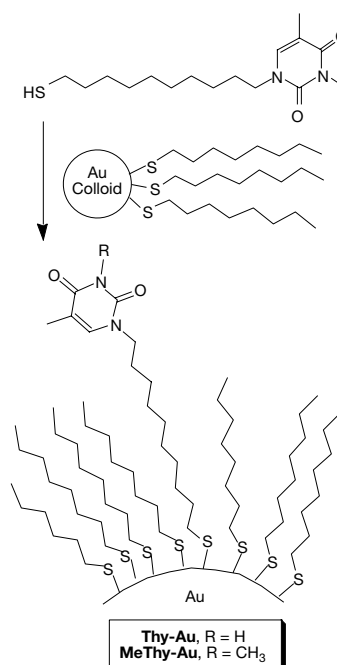


Figure 2 Preparation of Thy-Au and MeThy-Au.

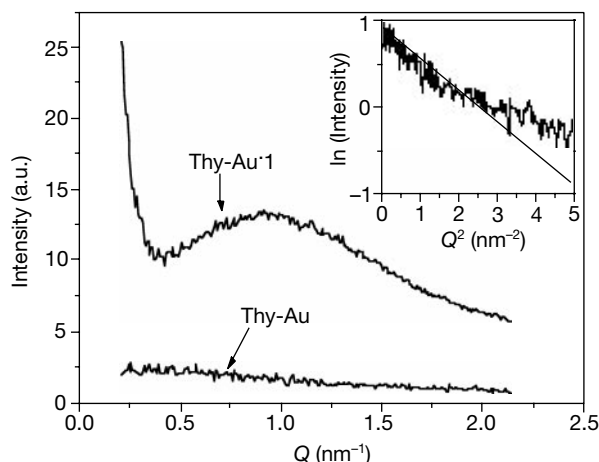


Figure 3 Small-angle X-ray scattering of Thy-Au and polymer **1**-Thy-Au aggregate. Q , scattering vector; a.u., arbitrary units. Inset shows a Guinier plot of Thy-Au in toluene, along with a fit of initial data yielding a radius of gyration of 1.3 nm.

Au particles *in situ* before and after this aggregation, using Ni-filtered Cu K α radiation. A Guinier plot of the data for the Thy-Au particles in toluene is shown in Fig. 3, inset, where the natural logarithm of the scattered intensity was plotted as a function of the square of the scattering vector, Q^2 ($Q = (4\pi/\lambda) \sin\theta$, where λ is the wavelength and 2θ is the scattering angle). The radius of gyration, obtained from the initial slope of the Guinier plot, was found to be 1 nm, corresponding to a weight average particle radius of 1.3 nm. This agrees with the average Au particle radius of 1.0 ± 0.3 nm as measured by transmission electron microscopy (TEM) (see below). After aggregation induced by polymer **1**, the SAXS plot of the precipitates shows marked changes (Fig. 3), exhibiting a distinct maximum at $Q = 0.94 \text{ nm}^{-1}$ characteristic of a defined separation distance between neighbouring Au particles. After background subtraction, the centre-to-centre distance of Au particles was determined to be 6.4 ± 0.3 nm in the polymer **1**-Thy-Au aggregate, corresponding to a 4.4-nm interparticle distance. A sharp increase in the scattering was observed at small Q , suggesting the presence of larger-scale structure (>20 nm) beyond instrumental resolution.

A model for the self-assembly observed in the aggregated structure is shown in Fig. 4a. In non-polar solvents, polymer **1** folds into a highly compact structure owing to intramolecular hydrogen bonds between the triazines²¹. Multivalent interactions of polymer **1** with Thy-Au induce the unfolding of the compact structure of **1**, exposing further triazine recognition units. This allows polymer **1** to interact with further Thy-Au units, propagating the assembly process.

Insight into the structure of the polymer **1**-Thy-Au aggregate can be obtained through molecular modelling (Fig. 4b). Using a configuration of the polymer chain that spans the gold particles, a particle-particle distance of 4.4 nm is obtained, agreeing very well with the distance of 4.4 ± 0.3 nm determined by SAXS. This provides a working model for the polymer conformation between the gold particles. The influence of polymer molecular mass, polymer functionalization, particle size and particle functionalization on the self-assembly process is currently under investigation.

Proof of the large-scale ordering suggested by SAXS was obtained using TEM (Fig. 5). Micrographs of the THF-soluble fraction of the polymer **1**-Thy-Au precipitate (Fig. 5a) reveal the formation of large, highly regular, spherical clusters with diameters of 97 ± 17 nm. These clusters consist of 3,000–7,000 individual nanoparticles per microsphere. We note that the self-assembled microspheres are stable in polar solvents known to disrupt hydrogen-bond interactions, providing further evidence that the macrostructure is

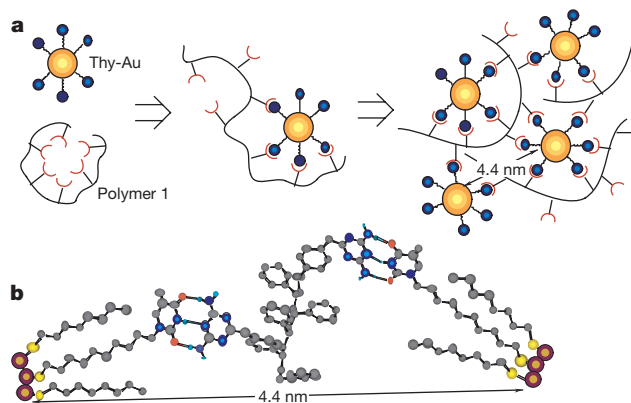


Figure 4 Proposed mechanism for the aggregation of polymer **1**-Thy-Au. **a**, Polymer mediated self-assembly of Thy-Au, showing the experimentally determined interparticle distance. **b**, Proposed polymer **1**-Thy-Au self-assembled structure (using the AMBER force field as implemented by MacroModel 4.0), showing the computationally predicted interparticle distance.

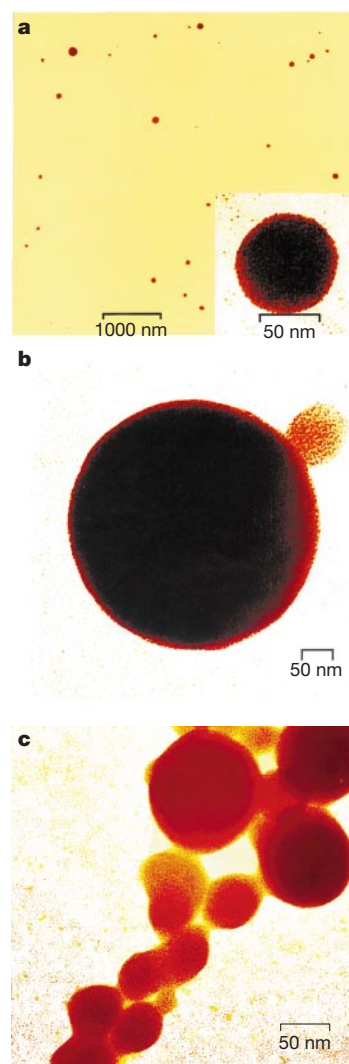


Figure 5 TEM image of polymer **1**-Thy-Au aggregates formed at different temperatures. **a**, 23 °C. Inset, representative self assembled microparticle. **b**, -20 °C. **c**, 10 °C.

stabilized by polyvalent polymer–colloid interactions. Roughly spherical aggregates of similar dimensions (~ 60 nm) were obtained from 6-nm analogues of Thy-Au and polymer **1** (see Supplementary Information), demonstrating the applicability of our polymer-mediated approach to the assembly of different sized nanoparticle subunits.

Self-assembly processes are governed by a balance of entropic and enthalpic effects, making them very temperature dependent. This temperature dependence is manifested by more efficient recognition processes at lower temperatures²¹, which would be expected to yield larger aggregate structures. Investigations of temperature effects on the preparation of the aggregates yielded results consistent with this prediction. TEM micrographs of the precipitate formed at -20°C revealed the formation of microscale ($0.5\text{--}1\ \mu\text{m}$) discrete spherical particles (Fig. 5b), consisting of $(6\text{--}50) \times 10^5$ individual Thy-Au units. These microscale particles are among the most complex synthetic self-assembled structures known, demonstrating the thermal control of aggregate size using the ‘bricks and mortar’ methodology.

In addition to controlling the size of the aggregates, temperature strongly affects the morphology of the resulting ensembles. At 10°C , networks were formed (Fig. 5c), as opposed to the discrete structures observed at higher and lower temperatures. This suggests that network formation is an intermediate process in the formation of the giant assemblies at -20°C . The individual assemblies within these networks remained spherical, although their sizes are more highly dispersed. We are currently investigating methods aimed at achieving control over the size and the geometry of these networks, which would allow the fabrication of rod- and wire-like structures for incorporation in nanoscale constructs (see Supplementary Information).

We have demonstrated a polymer-mediated strategy for the self-assembly of nanoparticles into structured ensembles. Discrete microspheres formed using this methodology are truly multi-scale constructs, highly ordered on the molecular, nanometre and micro-metre scale. This method has also been shown to generate network structures, which can serve as precursors to other shapes and sizes of nanoparticle assembly. The degree of ordering and the control of particle size and shape, coupled with the inherent modularity of the ‘bricks and mortar’ colloid-polymer self-assembly process, represent a powerful and general strategy for the creation of highly structured multifunctional materials and devices. □

Received 8 July 1999; accepted 10 February 2000.

1. Anelli, P. L., Spencer, N. & Stoddart, J. F. A molecular shuttle. *J. Am. Chem. Soc.* **113**, 5131–5133 (1991).
2. Otsuki, J. *et al.* Redox-responsive molecular switch for intramolecular energy transfer. *J. Am. Chem. Soc.* **119**, 7895–7896 (1997).
3. de Silva, A. P., Gunaratne, H. Q. N. & McCoy, C. P. Molecular photoionic and logic gates with bright fluorescence and “off-on” digital action. *J. Am. Chem. Soc.* **119**, 7891–7892 (1997).
4. Alivisatos, A. P. From molecules to materials: Current trends and future directions. *Adv. Mater.* **10**, 1297–1336 (1998).
5. Breen, T. L., Tien, J., Oliver, S. R. J., Hadzic, T. & Whitesides, G. M. Design and self-assembly of open, regular, 3D mesostructures. *Science* **284**, 948–951 (1999).
6. Lehn, J.-M. *Supramolecular Chemistry* (VCH, Weinheim, 1995).
7. Fyfe, M. C. T. & Stoddart, J. F. Synthetic supramolecular chemistry. *Acc. Chem. Res.* **30**, 393–401 (1997).
8. Rebek, J. Reversible encapsulation and its consequences in solution. *Acc. Chem. Res.* **32**, 278–286 (1999).
9. Stupp, S. I. *et al.* Supramolecular materials: Self-organized nanostructures. *Science* **276**, 384–389 (1997).
10. Kotera, M., Lehn, J.-M. & Vigneron, J.-P. Self-assembled supramolecular rigid rods. *J. Chem. Soc. Chem. Commun.* 197–199 (1994).
11. Edgecombe, B. D., Stein, J. A., Fréchet, J. M. J., Xu, Z. & Kramer, E. J. The role of polymer architecture in strengthening polymer-polymer interfaces: A comparison of graft, block, and random copolymers containing hydrogen-bonding moieties. *Macromolecules* **31**, 1292–1304 (1998).
12. Andres, R. P. *et al.* Self-assembly of a two-dimensional superlattice of molecularly linked metal clusters. *Science* **273**, 1690–1693 (1996).
13. Markovich, G. *et al.* Architectonic quantum dot solids. *Acc. Chem. Res.* **32**, 415–423 (1999).
14. Liu, J. *et al.* Cyclodextrin-modified gold nanospheres. Host-guest interactions at work to control colloidal properties. *J. Am. Chem. Soc.* **121**, 4304–4305 (1999).
15. Marinakos, S. M., Brousseau, L. C. III, Jones, A. & Feldheim, D. L. Template synthesis of one-

- dimensional Au, Au-poly(pyrrole), and poly(pyrrole) nanoparticle arrays. *Chem. Mater.* **10**, 1214–1219 (1998).
16. Whetten, R. L. *et al.* Crystal structures of molecular gold nanocrystal arrays. *Acc. Chem. Res.* **32**, 397–406 (1999).
17. Mucic, R. C., Storhoff, J. J., Mirkin, C. A. & Letsinger, R. L. DNA-directed synthesis of binary nanoparticle network materials. *J. Am. Chem. Soc.* **120**, 12674–12675 (1998).
18. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
19. Alivisatos, A. P. *et al.* Organization of ‘nanocrystal molecules’ using DNA. *Nature* **382**, 609–611 (1996).
20. Boal, A. K. & Rotello, V. M. Redox-modulated recognition of flavin by functionalized gold nanoparticles. *J. Am. Chem. Soc.* **121**, 4914–4915 (1999).
21. Deans, R., Ilhan, F. & Rotello, V. M. Recognition mediated unfolding of a self-assembled polymeric globule. *Macromolecules* **32**, 4956–4960 (1999).
22. Park, T. K., Schroeder, J. & Rebek, J. New molecular complements to imides. Complexation to thymine derivatives. *J. Am. Chem. Soc.* **113**, 5125–5127 (1991).
23. Muehldorf, A. V., Van Engen, D., Warner, J. C. & Hamilton, A. D. Aromatic aromatic interactions in molecular recognition: a family of artificial receptors for thymine that shows both face-to-face and edge-to-face interactions. *J. Am. Chem. Soc.* **110**, 6561–6562 (1988).
24. Brust, M., Walker, M., Bethell, D., Schiffrin, D. J. & Whyman, R. Synthesis of thiol-derivatized gold nanoparticles in a 2-phase liquid-liquid system. *J. Chem. Soc. Chem. Commun.* 801–802 (1994).
25. Ingram, R. S., Hostetler, M. J. & Murray, R. W. Poly-hetero-omega-functionalized alkanethiolate-stabilized gold cluster compounds. *J. Am. Chem. Soc.* **119**, 9175–9178 (1997).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) and as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the NSF V.M.R. also acknowledges support from the Petroleum Research Fund of the ACS, the Center for University Research in Polymers at the University of Massachusetts, the Alfred P. Sloan Foundation, Research Corporation, and the Camille and Henry Dreyfus Foundation.

Correspondence and requests for materials should be addressed to V.M.R. (e-mail: rotello@chem.umass.edu).

Subduction erosion along the Middle America convergent margin

C. R. Ranero & R. von Huene

GEOMAR, Wischhofstrasse 1-3, Kiel 24148, Germany

‘Subduction erosion’ has been invoked to explain material missing from some continents along convergent margins¹. It has been suggested that this form of tectonic erosion removes continental material at the front of the margin or along the underside of the upper (continental) plate^{2–4}. Frontal erosion is interpreted from disrupted topography at the base of a slope and is most evident in the wake of subducting seamounts^{5,6}. In contrast, structures resulting from erosion at the base of a continental plate are seldom recognized in seismic reflection images because such images typically have poor resolution at distances greater than ~ 5 km from the trench axis. Basal erosion from seamounts and ridges has been inferred^{7,8}, but few large subducted bodies—let alone the eroded base of the upper plate—are imaged convincingly. From seismic images we identify here two mechanisms of basal erosion: erosion by seamount tunnelling and removal of large rock lenses of a distending upper plate. Seismic cross-sections from Costa Rica to Nicaragua indicate that erosion may extend along much of the Middle America convergent margin.

From Costa Rica to Nicaragua the Pacific continental margin is formed by a wedge-shaped body of igneous rock (margin wedge), covered by 1–2 km of slope sediment and fronted by a small sediment prism^{9–11}. The facing ocean plate has a variable structure; there is thick crust at the Cocos ridge¹², the plate is $\sim 40\%$ covered by seamounts off central Costa Rica, and has a smooth morphology off the northwest Nicoya peninsula (Fig. 1). Off Nicaragua, the sub-

ducting ocean crust is thin (~ 5 km) and cut by many normal faults with ~ 0.5 -km displacement^{11,13,14}. Variations of crustal structure and seafloor morphology of the subducting plate correspond with regional and local changes in upper-plate thickness along the margin. On a regional scale, the ocean crust thins and deepens towards Nicaragua, whereas the corresponding continental margin wedge thickens (Fig. 2). Where the buoyant Cocos ridge subducts, the margin wedge is thinner than opposite the adjacent seamount segment (Fig. 2). Facing the seamount segment the margin is thinner than to the northwest and is thinnest at a large embayment off the southeast Nicoya peninsula (Figs 1 and 2). The margin wedge thickens off the northwest Nicoya peninsula where the featureless sea floor subducts, and is even thicker off Nicaragua, where deep sea floor and thin crust subduct.

The thin margin wedge (Fig. 2) and indented continental slope (Fig. 1) opposite the seamounts and the Cocos ridge suggest that repeated seamount subduction is an effective agent of upper-plate erosion. Opposite the seamount segment the slope has several grooves parallel to the convergence vector (G1 to G4, Fig. 1). The grooves mark the path of subducting seamounts, and the local uplift at the end of G1, G2 and G4 indicates that three seamounts are underthrust beneath the slope¹⁴. The grooves indicate missing upper-plate material associated with seamount subduction. Two upper-plate areas are eroded during seamount subduction. Seamount impact erodes the margin front producing re-entrant

structures at the base of the continental slope (for example, G1 and G4, Figs 1 and 3a). Landward, in the 30–40 km of the margin upslope of the frontal prism, the margin wedge remains semi-coherent as seamount underthrusting uplifts, fractures and thins the upper plate. The resulting rough margin-wedge upper surface has relief locally exceeding 0.5 km (Fig. 3a, b). Discontinuous strata are locally folded parallel to the surface (Fig. 3b), and undeformed younger strata smoothes this topography. The rough surface and a disrupted canyon system are roughly coincident. Seismic images from areas where seamounts have subducted (for example, Fig. 3a) and along the grooves of subducting seamounts¹⁴ indicate that only the upper part of the slope sediment slides at the oversteepened slope above subducting seamounts. Although some of the slumped sediment is missing, it is not enough to explain the grooves. The groove G3 penetrates the most into the margin (Fig. 1) where it is a deep trough collecting sediment delivered by shallow canyons (Fig. 3a). Although the groove is partially filled with recent sediment, the anomalous depth indicates missing material and erosion at the base of the upper plate.

Where seamounts no longer disrupt the front of the margin wedge but tunnel beneath it, we have imaged active basal erosion. Above a seamount about 2 km high, 0.5–0.7 km of rock from the base of a 2-km-thick margin wedge is missing (Fig. 4), which helps to explain the grooves and the thin margin wedge at the embayments. It has been proposed that uplift of the decollement over a

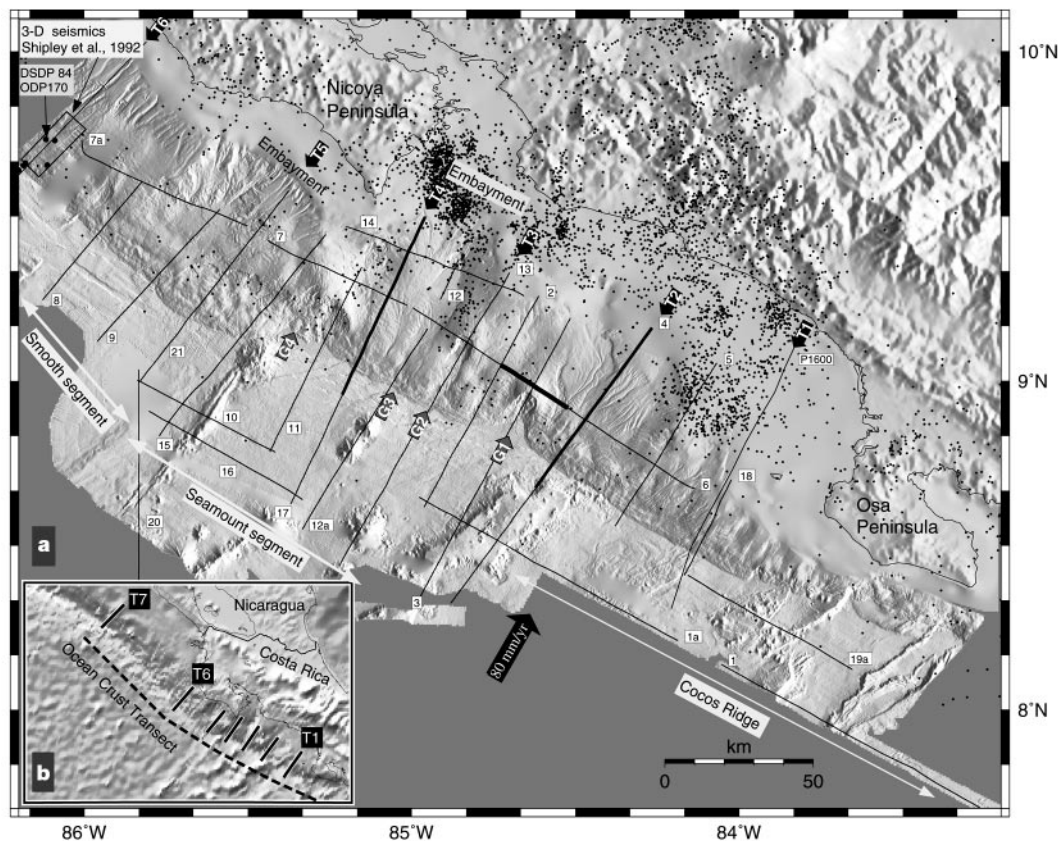


Figure 1 Bathymetry and tracks of seismic data off Costa Rica and Nicaragua. **a**, Multibeam hydrosweep bathymetry off Costa Rica with tracks of Sonne-81 and Shell P1600 multichannel seismic reflection data. The sudden change in roughness from the continental slope to the platform occurs where the multibeam bathymetry coverage ends. Bathymetry in the platform is from low-resolution maps. Thick lines mark segments shown in Figs 3 (lines 4 and 17) and 4, (line 6). Filled dots are earthquake epicenters¹⁷. Only earthquakes deeper than 30 km are plotted beneath the mainland. The ocean plate has three segments: the Cocos ridge with crust > 12 km thick¹², the seamount segment with topographic features 1–2.5 km tall and 10–20 km wide, and a smooth seafloor segment

off the northwest Nicoya peninsula. Subduction of seamounts is indicated by grooves in the continental slope (G1–G4) and clusters of earthquakes beneath the continental slope and shelf. Large embayments in the margin morphology indicate areas where seamount subduction has eroded the upper plate¹⁴. **b**, Location of the transect along the ocean plate (dashed line) and of the cross-sections T1–T7 used in Fig. 2. Cross-section T2 is a prestack depth migration. The other cross-sections are constrained by wide-angle and coincident near-vertical seismic data. Box off northwest Nicoya peninsula marks the area of a 3D seismic data set¹⁹.

seamount might create a fluid pressure gradient that drains fluid from the seamount flanks and concentrates it at the crest, thereby promoting hydrofracturing and low friction⁸. However, the pervasive fracturing observed in the bathymetry at the crest of domes above subducting seamounts (Fig. 3a) provides fluids conduits for venting that probably relieve pressure. The increase in normal stress across the plate interface resulting from the accommodation of the seamount¹⁵ in addition to the roughness of the volcano probably facilitates abrasion along the seamount subduction path¹⁶. Earthquakes of magnitude greater than 4 at the front of the margin¹⁷ are located near the subducting seamounts and have strike-slip rather than thrust focal mechanisms, indicating a relatively high friction between seamounts and upper plate. Where the upper plate is >6–8 km thick, seamounts uplift but do not break up the margin wedge (Fig. 3a), so its upper surface is smooth, disrupted only by landward dipping normal faults that extend into the overlying strata (Fig. 3c).

On several seismic lines, plate-boundary reflections deeper than ~6 km bifurcate and surround megalenses (10–15 km long, 1.5–2 km high) whose origin is puzzling (Fig. 3a and d). Similar structures are observed off the northwest Nicoya peninsula^{18,19}. The megalenses occur in areas where the upper plate extends by normal faulting, and could represent material transferred from the lower to the upper plate (for example, underplated sediment or a sheared-off seamount) or alternatively margin-wedge rock transferred to the lower plate. The size in two dimensions of some megalenses is equivalent to a seamount (Fig. 3d), so if they had been transferred from lower to upper plate the uplift should resemble that of currently underthrusting seamounts. However, seismics and bathymetry (Fig. 3d) show no large doming, but rather a relatively stable slope with a canyon system, and also a less disrupted margin wedge and slope sediment than along seamount paths, inconsistent with a subducting or sheared-off seamount. Underplating might require more time and thus the slope may have stabilized, but the ~2-km-thick lens would require a large uplift of the margin wedge (see, for example, Fig. 4) which is not observed (Fig. 3d). Thus, the megalenses are probably being removed from the upper plate. Further support for erosion comes from the style of normal faulting. Some reflections from the landward-dipping fabric of events within the margin wedge project to offsets at the top, indicating that normal faults cut deep into the upper plate. Landward-dipping, deep-penetrating normal faulting has also been described in the middle slope off northwest Nicoya¹⁸. On line 4, fault dip decreases and fault throw grows trenchwards

(Fig. 3e), with the largest fault throws seaward of the megalenses where the margin wedge is thinner (Fig. 3d). Extension, thinning and collapse of the margin seaward of the megalenses is the response to the material transfer. The megalenses occur at depths at which temperatures calculated from depth to bottom-simulating reflectors indicate that the smectite–illite transition has occurred, and the plate interface is entering the zone of stick-slip behaviour²⁰. Thus, the lenses could be associated with active erosion in an environment of increasing friction. This mechanism has no apparent relation to seamount subduction, and might be active in other areas where relatively smooth ocean floor is subducted. Beneath the upper slope, the upper boundary of the megalenses is a linear splay from the plate-boundary interface at about 10–11 km depth. The landward-dipping upper-boundary reflections have a similar or slightly shallower dip than linear reflections above, within the margin wedge, which might indicate that megalenses form using pre-existing planes of weakness. Movement along a pre-existing weak interface may be favoured when fluids expelled from subducting sediment reach the structure and reduce friction, which detaches a lens of margin wedge rock. Seaward, beneath the middle slope, the landward-dipping upper boundary of the megalenses merges with a subhorizontal group of reflections (Fig. 3d) which might indicate a lithological boundary within the margin wedge.

Tectonic erosion has been inferred from long-term subsidence of the continental slope off Nicaragua, and the steep slope with scars of numerous landslides indicates that erosion is probably currently active^{11,14}. Thus, tectonic erosion may not be limited to the region where thick crust and rough sea floor subducts, but could extend along much of the Middle America trench. However, thinning and deepening of the ocean crust, and the smaller seafloor relief towards Nicaragua, correspond to changes in tectonism along the plate boundary. Closely spaced normal faulting of the ocean plate off Nicaragua creates a ‘washboard’ topography. Subduction of this type of morphology has been proposed as an erosional mechanism²¹. However, the upper plate off Nicaragua is thick relative to that of northwest Nicoya; this indicates that the ‘washboard’ topography is either not an efficient mechanism of basal erosion or that erosion there is limited to the frontal ~10 km of the upper plate.

In addition to the erosional mechanisms discussed above, the subduction of the entire ocean pelagic-hemipelagic section indicates

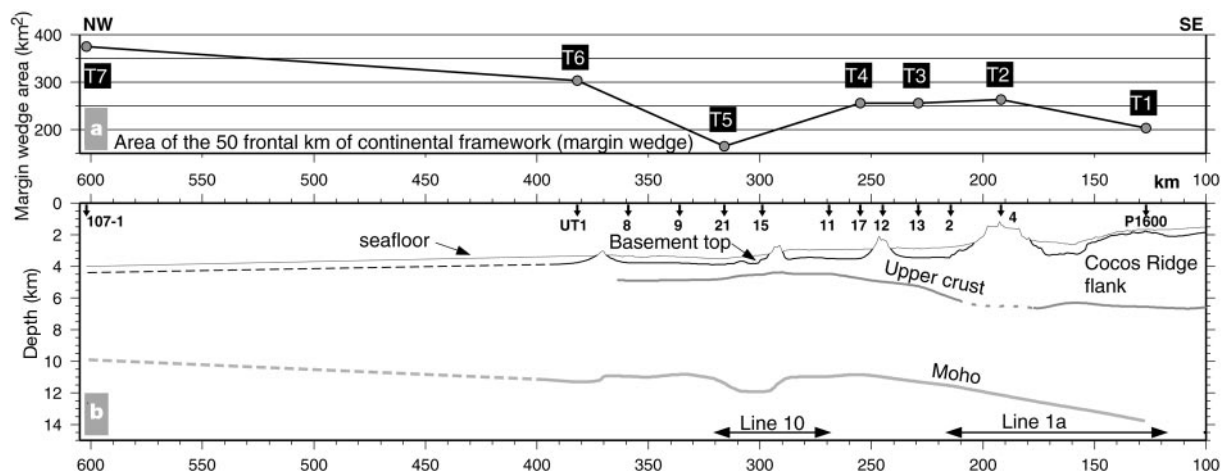


Figure 2 Comparison of continental framework thickness with the structure of the facing ocean plate. **a**, Cross-sectional area (km²) of the continental framework (margin wedge) of the upper plate from Costa Rica to Nicaragua. Values T1–T7 are the area of the frontal 50 km of margin wedge calculated along the transects shown in Fig. 1. **b**, Crustal

structure of the ocean plate along the transect in Fig. 1b. Downward-pointing arrows with numbers indicate crosspoints of transect with seismic tracks used to constrain the crustal structure. Double-headed arrows indicate extent of seismic lines along the transect. Data from refs 12, 13, 23 and this study.

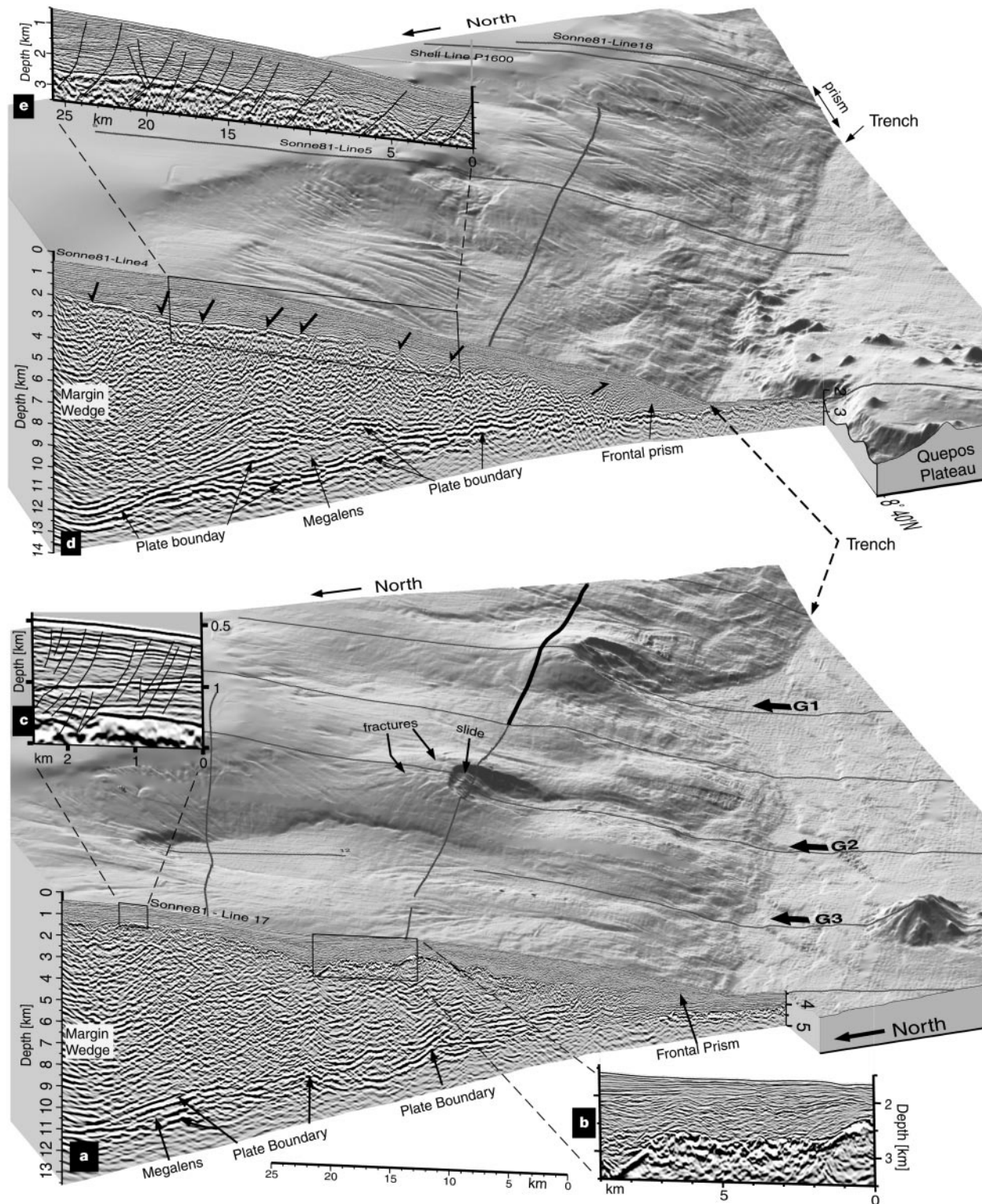


Figure 3 Prestack depth migration of Sonne-81 lines 17 and 4 projected on bathymetry perspectives. The iterative migration procedure uses velocities constrained by focusing analyses and common reflection point gathers²⁴. Wide-angle data guided velocity picking at depths greater than ~5 km. Resolution at ~10 km is ~0.5 km. **a**, Line 17 is 55 km long from the toe to the upper continental slope. The line is where seamounts have recently subducted. Note the small frontal prism, the continuity of plate boundary reflections and the distinct top of the margin wedge. The frontal ~40 km of the margin have a rough margin-wedge top (inset **b**) produced by underthrusting of seamounts like those shown in the bathymetry. Arrows point to grooves in the bathymetry (G1–G3) produced by subducting seamounts, and the doming up of the sea floor indicates the location of two of those seamounts. Where the margin wedge is >6–8 km thick its top is

smooth, only cut by normal faulting (inset **c**). Beneath the upper slope, plate-boundary reflections bifurcate and surround a rock megalens. **d**, Line 4 is 58 km long from the toe to the upper continental slope. The top of the margin wedge is smooth beneath the upper slope, cut by small throw normal faults (inset **e**). Fault throw grows and fault dip decreases downslope and the top of the margin wedge becomes rougher. The canyons in the upper and middle slope indicate a relatively stable environment where no seamount has recently underthrust. Thrust faulting occurs only at the small frontal prism. The frontal prism is a distinct morphological unit, tectonically active as indicated by the disruption of the slope drainage system. Plate-boundary reflections bifurcate at ~6 km depth and surround a rock megalens.

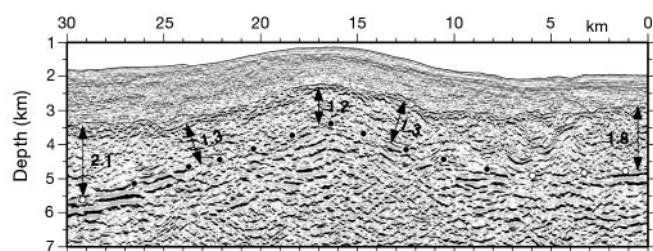


Figure 4 Active tectonic erosion by seamount tunnelling. The seismic image is a prestack depth migration of a segment of Sonne-81 line 6 over a subducting seamount. Dots mark the plate boundary, with black dots delineating the seamount flanks. This segment of the seismic line is parallel to the continental margin, and shows lateral variations in margin wedge thickness. The margin wedge is 0.5–0.7 km thinner above the subducting seamount. Extension of the upper plate by uplift above the seamount is too small to explain the thinning. Thus, thinning is probably due to tectonic erosion produced by mechanical wearing of the upper plate. The location of this cross-section is shown on Figs 1 and 3.

that hydrofracturing and piecemeal stoping^{2,4,8} (an underground mining process) of the base of the upper plate by overpressured fluids may also contribute to upper-plate thinning. A body of Plio-Pleistocene sediment at the front of the margin from southern Costa Rica to Guatemala is limited to a prism less than 10 km wide^{11,14,22}. Although underplating at the rear of the frontal prism might be active or an older relatively small body of accreted material may be present¹⁹, accretion of sediment must be limited in space and time. Subduction erosion at present dominates processes off Costa Rica, and probably also extends into the Nicaraguan margin. □

Received 6 August 1999; accepted 22 February 2000.

- Rutland, R. W. R. Andean orogeny and ocean floor spreading. *Nature* **233**, 252–255 (1971).
- Murauchi, S. & Ludwig, W. J. Crustal structure of the Japan trench: The effect of subduction of ocean crust. *Init. Rep. DSDP 56-57* (part 1), 463–470 (1980).
- Scholl, D. W., von Huene, R., Vallier, T. L. & Howell, D. G. Sedimentary masses and concepts about tectonic processes at underthrust ocean margins. *Geology* **8**, 564–568 (1980).
- von Huene, R. & Scholl, D. W. Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust. *Rev. Geophys.* **29**, 279–316 (1991).
- Lallemant, S. & Le Pichon, X. Coulomb wedge model applied to the subduction of seamounts in the Japan Trench. *Geology* **15**, 1065–1069 (1987).
- Collot, J. Y. & Fisher, M. A. Formation of forearc basins by collision between seamounts and accretionary wedges: An example from the New Hebrides subduction zone. *Geology* **17**, 939–933 (1989).
- Ballance, P. F. *et al.* Subduction of a large Cretaceous seamount of the Louisville ridge at the Tonga Trench: A model of normal and accelerated tectonic erosion. *Tectonics* **8**, 953–962 (1989).
- Le Pichon, X., Henry, P. & Lallemant, S. Accretion and erosion in subduction zones: The role of fluids. *Annu. Rev. Earth Planet. Sci.* **21**, 307–331 (1993).
- Hinz, K., von Huene, R., Ranero, C. R. & PACOMAR working group. Tectonic structure of the convergent Pacific margin offshore Costa Rica from multichannel seismic reflection data. *Tectonics* **15**, 54–66 (1996).
- Kimura, G. *et al.* *Init. Rep. ODP 170* (1997).
- Ranero, C. R. *et al.* A cross section of the Pacific Margin of Nicaragua. *Tectonics* (in the press).
- Stavenhagen, A. U. *et al.* Seismic wide-angle investigations in Costa Rica—a crustal velocity model from the Pacific to the Caribbean Coast. *Zbl. Geol. Paläontol.* **1**, 393–408 (1998).
- Walther, C. H. E., Flueh, E. R., Ranero, C. R., von Huene, R. & Strauch, W. Crustal structure across the Pacific Margin of Nicaragua—evidence for ophiolitic basement and a shallow mantle sliver. *Geophys. J. Int.* (in the press).
- von Huene, R., Ranero, C. R., Weinrebe, W. & Hinz, K. Quaternary convergent margin tectonics of Costa Rica, segmentation of the Cocos Plate, and Central American volcanism. *Tectonics* (in the press).
- Scholz, C. H. & Small, C. The effect of seamount subduction on seismic coupling. *Geology* **25**, 487–490 (1997).
- Wang, W. & Scholz, C. H. Wear processes during frictional sliding of rock: A theoretical and experimental study. *J. Geophys. Res.* **99**, 6789–6799 (1994).
- Protti, M., Güendel, F. & McNally, K. In *Geologic and Tectonic Development of the Caribbean Plate Boundary in Southern Central America* (ed. Mann, P.) (Special Paper 295, Geological Society of America, Boulder, Colorado, 1995).
- McIntosh, K., Silver, E. & Shipley, T. Evidence and mechanisms for forearc extension at the accretionary Costa Rica convergent margin. *Tectonics* **12**, 1380–1392 (1993).
- Shipley, T. H., McIntosh, K. D., Silver, E. A. & Stoffa, P. L. Three-dimensional seismic imaging of the Costa Rica accretionary prism: Structural diversity in a small volume of the lower slope. *J. Geophys. Res.* **97**, 4439–4459 (1992).
- Vrolijk, P. On the mechanical role of smectite in subduction zones. *Geology* **18**, 703–707 (1990).
- Hilde, T. W. C. Sediment subduction versus accretion around the Pacific. *Tectonophysics* **99**, 381–397 (1983).

- Aubouin, J. *et al.* Leg 84 of the deep drilling project: subduction without accretion: Middle America Trench off Guatemala. *Nature* **297**, 458–460 (1982).
- Ye, S. *et al.* Crustal structure of the subduction zone off Costa Rica derived from OBS refraction and wide-angle reflection seismic data. *Tectonics* **15**, 1006–1021 (1996).
- Mackay, S. & Abma, R. Depth focusing analysis using wavefront-curvature criterion. *Geophysics* **58**, 1148–1156 (1993).

Acknowledgements

We thank T. Reston for comments on the manuscript. SONNE-81 seismic reflection data were collected by the BGR.

Correspondence and requests for materials should be addressed to C.R.R. (e-mail: cranero@geomar.de).

Quantitative evidence for global amphibian population declines

Jeff. E. Houlahan*, C. Scott Findlay*†, Benedikt R. Schmidt‡, Andrea H. Meyer§ & Sergius L. Kuzmin||

* Ottawa-Carleton Institute of Biology, University of Ottawa, 30 Marie Curie, Ottawa, Ontario K1N 6N5, Canada

† Institute of Environment, University of Ottawa, 555 King Edward Street, Ottawa, Ontario K1N 6N5, Canada

‡ Zoologisches Institut, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

§ Swiss Federal Statistical Office, Sektion Hochschulen und Wissenschaft, Espace de l'Europe 10, CH-2010, Neuchâtel, Switzerland

|| Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow 117071, Russia

Although there is growing concern that amphibian populations are declining globally^{1–3}, much of the supporting evidence is either anecdotal^{4,5} or derived from short-term studies at small geographical scales^{6–8}. This raises questions not only about the difficulty of detecting temporal trends in populations which are notoriously variable^{9,10}, but also about the validity of inferring global trends from local or regional studies^{11,12}. Here we use data from 936 populations to assess large-scale temporal and spatial variations in amphibian population trends. On a global scale, our results indicate relatively rapid declines from the late 1950s/early 1960s to the late 1960s, followed by a reduced rate of decline to the present. Amphibian population trends during the 1960s were negative in western Europe (including the United Kingdom) and North America, but only the latter populations showed declines from the 1970s to the late 1990s. These results suggest that while large-scale trends show considerable geographical and temporal variability, amphibian populations are in fact declining—and that this decline has been happening for several decades.

Over the past several decades, there have been reports of catastrophic declines and extirpations of amphibian species in Australia, South and Central America^{13–16}, and in high-altitude regions of the western United States¹⁷. Ancillary evidence has pointed to several possible underlying causes, including changes in local climate¹⁸, acid precipitation¹⁹, disease^{20,21}, increased UV-B irradiation²² or various combinations thereof^{23,24}. Because these reports usually document the disappearances of species from small geographical regions, extrapolations to 'global' amphibian declines are tenuous at best. To address this problem of restricted geographical coverage, we have analysed 936 amphibian population data sets collected from journal publications and technical reports, as well as unpublished data provided by herpetologists from around the world. Every effort was made to be exhaustive, and we believe

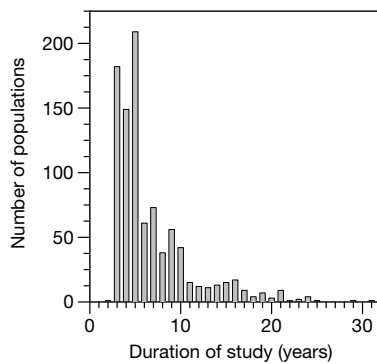


Figure 1 Frequency distribution of study durations.

that our sample represents the most complete collection of amphibian population time series to date. In total, over 200 researchers have contributed data from 37 countries and 8 regions of the world, including population data for 157 species from 21 families (Table 1). The studies range from 2 to 31 years in duration (Fig. 1).

Using a procedure (ΔN , see Methods) that we developed to detect population trends across multiple populations, and considering all 936 populations together (referred to here as the 'global' set), we find significant declines from approximately 1960 to the present (Fig. 2). There was no significant trend from 1950 to 1960 ($\beta = -0.003 \pm 0.013$, $t = 0.23$, $P = 0.50$), a steep decline from 1960 to 1966 ($\beta = -0.070 \pm 0.018$, $t = -3.96$, $P < 0.001$) followed by a shallower decline from 1966 to 1997 ($\beta = -0.008 \pm 0.001$, $t = -11.49$, $P < 0.001$). The trends from 1960–1966 and 1966–1997 represent annual declines of approximately 15% and 2%, respectively (Fig. 2). This apparent difference in rates of decline for the two time periods should be treated cautiously because of the relatively small sample sizes for the early years (Fig. 2).

To address potential biases introduced by the large number of short time series and variable data quality, we have in addition analysed trends in three non-independent subsets of the data, as described in the Methods section. These additional analyses are available as Supplementary Information. Although there are quantitative differences, the population patterns for these three subsets are qualitatively similar to the patterns found in the 'global' set (see Figs 1, 3, 5 in Appendix 2 of Supplementary Information).

Separate analyses of population data from North America and western Europe suggest that, while both showed declines between 1960 and the late 1960s, only in North America did populations appear to decline during the period 1966–1997 (North America, 1960–1997: $\beta = -0.026 \pm 0.002$, $t = -16.79$, $P < 0.001$; western

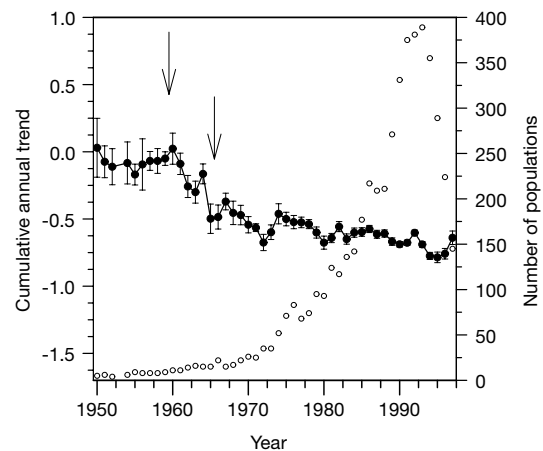


Figure 2 Amphibian population trends from 1950 to 1997 using all 936 populations. (Filled circles, $\Sigma \Delta N$; open circles, number of populations used to calculate ΔN). Arrows indicate the 'switchpoints' (see text).

Europe, 1960–66: $\beta = -0.158 \pm 0.027$, $t = -5.96$, $P < 0.001$; western Europe, 1966–1997: $\beta = 0.001 \pm 0.001$, $t = 1.18$, $0.2 < P < 0.4$) (Fig. 3a, b). The decline from 1960 to 1966 was much more pronounced in western Europe than in North America, but, because the negative trend has persisted over the past several decades in North America, the cumulative change in population size from 1960 to 1997 is similar for both regions. Results for the three subsets, again, show qualitatively similar patterns to analyses presented here despite minor quantitative differences (see Figs 2, 4, 6 in Appendix 2 of Supplementary Information). Regional analyses of low-latitude areas such as South America, Africa and Australia were not possible with this method because most of the populations are from North America and Europe (areas of low to intermediate amphibian diversity). Although there is increasing awareness of the need for more research on amphibian population dynamics²⁵, one clear message of our research is the pressing need for more studies from high-diversity regions of the world, such as the tropics.

A second method for assessing declines based on the ratio of declining to increasing populations (see Methods) indicates that, across all regions included in the analysis, there were significantly more declines than increases (56.6% declines, $\chi^2 = 15.94$, d.f. = 1, $P < 0.001$) (Table 2). However, there was no statistically significant difference in the ratio of declining to increasing populations among regions ($\chi^2 = 4.12$, d.f. = 4, $P = 0.39$). The data in the Supplementary Information also show significantly more declines than increases across all included regions. However, the peer-reviewed/published

Table 1 Summary information of 936 amphibian populations

Region (no. of populations, species and families)	Study duration Mean $\pm$ 1 s.d. (range)	1940–49	1950–59	1960–69	1970–79	1980–89	1990–98
W. Europe (511, 21, 7)	6.9 $\pm$ 4.2 (3–29)	0	0	7	45	171	419
N. America (240, 75, 12)	6.0 $\pm$ 4.1 (3–20)	2	6	15	61	120	110
UK (75, 6, 3)	10.3 $\pm$ 5.3, (3–25)	0	6	6	13	59	59
S./C. America (51, 34, 4)	4.7 $\pm$ 2.2 (3–15)	0	0	0	0	23	31
Australia/NZ (24, 8, 3)	3.9 $\pm$ 1.2 (3–6)	0	0	0	1	15	8
Asia (21, 12, 5)	5.7 $\pm$ 5.7 (3–22)	0	1	8	5	9	5
E. Europe (9, 6, 3)	9.9 $\pm$ 8.4 (4–31)	1	1	3	3	5	3
Africa/Middle East (5, 4, 4)	53.5 (2–11)	0	1	1	1	3	0

Shown are number of studies and study duration for each decade and geographical region. Populations contribute to more than one decade total if they were studied over more than one decade.

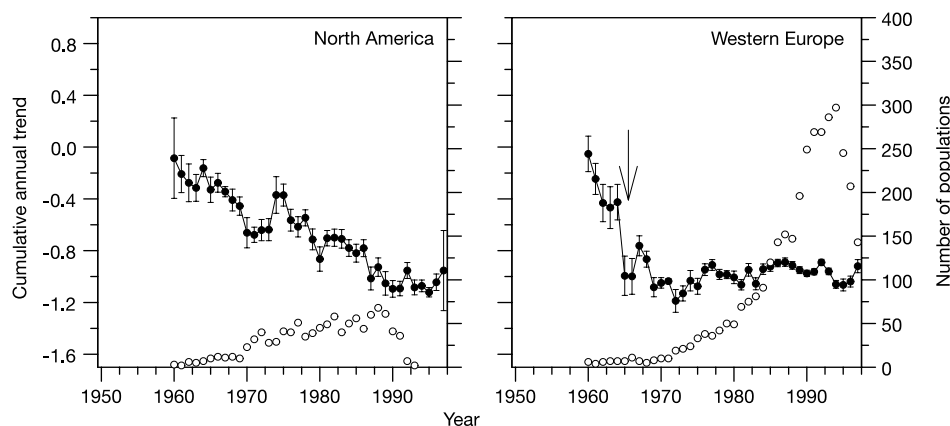


Figure 3 Amphibian population trends from 1960 to 1997 for North America and western Europe (including the United Kingdom). (Filled circles, $\Sigma \Delta N$; open circles, number of populations used to calculate ΔN). Arrows indicate the 'switchpoints' (see text).

and the published data sets do show significant differences in the ratio of declining to increasing populations among regions (see Tables 1–6 in Appendix 3 in Supplementary Information). An examination of the standardized deviates of the log-linear models for the published and peer-reviewed/published subsets shows that North America and Australia/New Zealand have relatively more declines than other regions.

Our analysis of amphibian population trends from around the world suggests that while there is considerable geographical and temporal variability, at a global scale amphibians have declined over the past several decades and continue to do so. Most of the evidence to date for global amphibian declines has involved documentation of local extirpations of populations^{5,18}. Only 61 of the 936 populations that we studied went extinct (without subsequent recolonization during the study), and only one of those was during the period of most significant decline (1960–66), providing evidence that even extant populations have undergone strong declines. Our analysis

further suggests recent declines, especially in North America and Australia/New Zealand, two regions where anecdotal evidence of declines is perhaps the strongest^{26,27}. We also found evidence for historical declines in both Europe and North America. Indeed, our results suggest that the most dramatic declines—from which amphibian populations and communities seem not to have recovered—occurred several decades before herpetologists sounded the alarm. □

Methods

Data description and selection

Data from 936 amphibian populations were used in this study (for supporting documentation, see Appendix 1 of Supplementary Information). In combining results from many studies, there must be explicit criteria for deciding which studies to include. Recent work on meta-analytic techniques suggests that the more restrictive the criteria for inclusion, the greater the potential bias²⁸. On the other hand, some attempt must be made to evaluate the robustness of the results to changes in inclusion criteria (see Table 3). To address this issue, we analysed the 'global' set and three different subsets of the 'global' set defined by (1) the duration of the population time-series; (2) whether the data were published; and (3) whether the published data had been peer reviewed. Here we present the results of our analysis of the 'global' set; all results for analyses using the three subsets (that is, data sets with more restrictive criteria) are presented in Appendices 1–3 in Supplementary Information.

For all population data, estimates given as a range (for example, 35–40 individuals) were converted to the mean of the upper and lower values of the range. Where estimates were a minimum (for example, 50+), we have taken the minimum estimate as the population estimate (this occurred in fewer than ten populations). Some populations were counted more than once a year. For breeding aggregations, the maximum abundance estimate was used, and for non-breeding populations the average of all abundance estimates in a calendar year was used.

Analysis

Two types of analyses were used: we called one the 'ΔN method' and the other 'the proportion declining' method. The former method combines measurements of

Table 2 Individual population trends by region

Region	No. of declining populations	No. of increasing populations	No. of no-trend populations
Western Europe	309	248	29
North America	130	96	14
South America	31	19	1
Australia/New Zealand	17	6	1
Asia	10	10	1
Eastern Europe	4	5	0
Africa/Middle East	2	2	1

Trends of individual populations were calculated using Kendall's τ , Pearson correlation and Spearman rank correlation. All three techniques gave qualitatively the same result, so only Kendall's τ results are presented here.

Table 3 Selection criteria

Selection criteria	Data Set			
	'Global'	Subset 1	Subset 2	Subset 3
Population or relative abundance estimates were available for at least two years	X		X	X
No change in methodology*	X	X	X	X
No experimental manipulation	X	X	X	X
At least 10 animals counted in at least 1 year of the study	X	X	X	X
No categorical measures†	X	X	X	X
Study initiated since 1950	X	X	X	X
Estimates were available for at least seven years		X		
Published‡			X	
Peer-reviewed§				X

* Studies with changes that would lead to obvious biases in population trends, for example, trends over time in search/capture effort, were excluded. Techniques for taking censuses of amphibian populations included egg mass counts, visual encounter surveys, quadrat and transect sampling, drift fence/pit trap counts, removal sampling, mark-recapture estimates, calling male counts and aquatic traps³⁰.

† Abundance estimates were derived from integer counts of individuals, egg masses, or vocalizations.

‡ Data had to have been published in peer-reviewed journals or in government or technical publications.

§ Data had to have been published in a peer-reviewed journal. Where there was any doubt as to whether a publication was peer-reviewed, the data were excluded from the analyses.

population change across multiple populations, and asks how large the 'average' change in population size is over time; the latter method simply measures trends in individual populations, and asks whether the ratio of declining populations to increasing populations is significantly different from 1:1. The first method is preferable, because it allows us to assess potential changes in the rate of decline over time, but it requires large sample sizes that were not available for some regions. The second method allowed us to test for declines in those regions with small sample sizes.

ΔN method

This method was used to test for trends in the 'global' set, and to test for 'regional' trends in North America and western Europe (including the UK). For 'global' and 'regional' trends we calculated $\log(N+1)_{t+1} - \log(N+1)_t = \Delta N$ for successive yearly intervals. For this analysis, only populations having at least two consecutive years of data could be used. We then calculated $\Delta N = (\sum_{t=1}^n \Delta N_t)/n$ based on all populations (n) for which there were data for the time interval ($t, t+1$) in question (because the number of studies increases over time, so does the sample size for ΔN). This procedure was repeated for each year from 1950 to 1997, and the annual averages used to compute the cumulative average change, $\Delta N = \sum_{t=1}^T \Delta N_t$, from $T = 1950$ to $T = 1997$. More than 200 western European population time series come from two large studies in Sweden and Switzerland. Analyses of population trends with and without these data indicate they had no qualitative effect on the results.

Visual examination of a plot of ΔN for the global set suggests three qualitatively different time periods, corresponding roughly to 1950–1960, 1960–1970 and 1970–1997. To estimate switchpoints between periods, we fitted regression models including dummy (categorical) variables defining the period intervals (for example, period 1: 1950–1960; period 2: 1961–1970, and so on). Changing the beginning and end points for a given period results in a change in model fit, with the best estimate of the switchpoints derived from the model with the lowest residual mean square. The initial switchpoints were selected by examination of the data, with subsequent fitting based on moving the switchpoints forwards and backwards from the initial estimate(s). Model fitting ended when models with switchpoints two years earlier and later than the best model had higher residual mean-square values. Our best-fit model partitioned the global set into three time periods: 1950–1960, 1960–1966 and 1966–1997. The best model for the western European data showed two distinct time periods (1960–1966 and 1966–1997), while the best model for North America showed a single trend from 1960 to 1997.

ΔN s are summary data. As such, using it as the dependent variable in regression underestimates the true error sums of squares. We have corrected for this by including the error contribution of each ΔN to ΔN (ref. 29), and all significance tests use this true error sum of squares and the corresponding true degrees of freedom.

Proportion of declining populations method

In a second analysis, we evaluated trends in population size over time using the Spearman correlation, the Pearson correlation coefficient and Kendall's τ . Irrespective of the test statistic used, results were qualitatively the same. We present our results using Kendall's τ because it avoids some of the assumptions about data distribution. We calculated the correlation (Kendall's τ) between population size and year for each population in a particular geographical region. Populations were then classified as to whether they were declining (negative correlation), increasing (positive correlation) or had no trend (correlation = 0). For a population to show no trend, the correlation must be exactly 0. A log-linear model was fitted using the independent variables region, trend and their interaction (region $\times$ trend). Populations showing no trend were not included in the log-linear model. For the 'global' dataset and the three subsets, the Eastern European and African/Middle Eastern regions are presented but not included in the log-linear model; for the ≥ 7 year data set, Asia, South America and Australia/New Zealand are also presented, but excluded from the log-linear model because of low sample sizes.

Received 19 October 1999; accepted 26 January 2000.

- Wake, D. B. Declining amphibian populations. *Science* **253**, 860 (1991).
- Blaustein, A. R. & Wake, D. B. Declining amphibian populations: A global phenomenon? *Trends Ecol. Evol.* **5**, 203–204 (1990).
- Blaustein, A. R., Wake, D. B. & Sousa, W. P. Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. *Conserv. Biol.* **8**, 60–71 (1994).
- Wake, D. B. & Morowitz, H. J. Declining amphibian populations—a global phenomenon? Findings and recommendations. *Alytes* **9**, 33–42 (1991).
- Pechmann, J. H. K. & Wilbur, H. M. Putting declining amphibian populations in perspective: natural fluctuations and human impacts. *Herpetologica* **50**, 65–84 (1994).
- Lips, K. R. Decline of a tropical montane amphibian fauna. *Conserv. Biol.* **12**, 106–117 (1998).
- Harte, J. & Hoffman, E. Possible effects of acidic deposition on a Rocky Mountain population of the tiger salamander *Ambystoma tigrinum*. *Conserv. Biol.* **3**, 149–158 (1989).
- Corn, P. S. & Fogleman, J. C. Extinction of montane populations of the northern leopard frog (*Rana pipiens*) in Colorado. *J. Herpetol.* **18**, 147–152 (1984).
- Pechmann, J. H. K. *et al.* Declining amphibian populations: The problem of separating human impacts from natural fluctuations. *Science* **253**, 892–895 (1991).
- Meyer, A. H., Schmidt, B. R. & Grossenbacher, K. Analysis of three amphibian populations with quarter-century long time-series. *Proc. R. Soc. Lond. B* **265**, 523–528 (1998).
- Hecnar, S. J. & McCloskey, R. T. Spatial scale and determination of species status of the green frog. *Conserv. Biol.* **11**, 670–682 (1997).
- Hecnar, S. J. & McCloskey, R. T. Regional dynamics and the status of amphibians. *Ecology* **77**, 2091–2097 (1996).
- Weygoldt, P. Changes in the composition of mountain stream frog communities in the Atlantic mountains of Brazil—frogs as indicators of environmental deteriorations. *Stud. Neotrop. Fauna Environ.* **24**, 249–255 (1989).

- Berger, L. *et al.* Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
- Pounds, J. A., Fogden, M. P. L., Savage, J. M. & Gorman, G. C. Tests of null models for amphibian declines on a tropical mountain. *Conserv. Biol.* **11**, 1307–1322 (1997).
- Laurance, W. F., McDonald, K. R. & Speare, R. Epidemic disease and the catastrophic decline of Australian rain forest frogs. *Conserv. Biol.* **10**, 406–413 (1996).
- Drost, C. A. & Fellers, G. M. Collapse of a regional frog fauna in the Yosemite area of the California Sierra Nevada, USA. *Conserv. Biol.* **10**, 414–425 (1996).
- Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
- Beebe, T. J. C. *et al.* Decline of the natterjack toad *Bufo calamita* in Britain: paleoecological, documentary and experimental evidence for breeding site acidification. *Biol. Conserv.* **37**, 59–71 (1990).
- Carey, C. Hypothesis concerning the causes of the disappearance of boreal toads from the mountains of Colorado. *Conserv. Biol.* **7**, 355–362. (1993).
- Cunningham, A. A. *et al.* Unusual mortality associated with poxvirus-like particles in frogs (*Rana temporaria*). *Vet. Rec.* **133**, 141–142 (1993).
- Blaustein, A. R. *et al.* Effects of ultraviolet radiation on amphibians: Field experiments. *Am. Zool.* **38**, 799–812 (1998).
- Kiesecker, J. M. & Blaustein, A. R. Synergism between UV-B radiation and a pathogen magnifies amphibian embryo mortality in nature. *Proc. Natl Acad. Sci. USA* **92**, 11049–11052 (1995).
- Long, L. E., Saylor, L. S. & Soule, M. E. A pH/UV-B synergism in amphibians. *Conserv. Biol.* **9**, 1301–1303 (1995).
- Wake, D. B. Action on amphibians. *Trends Ecol. Evol.* **13**, 379–380 (1998).
- Tyler, M. J. Declining amphibian populations: a global phenomenon? An Australian perspective. *Alytes* **9**, 43–50 (1990).
- Green, D. M. in *Amphibians in Decline: Canadian Studies of a Global Problem* (ed. Green, D. M.) 291–308 (Society for the Study of Amphibians and Reptiles, Lawrence, USA, 1997).
- Englund, G., Sarnelle, O. & Cooper, S. D. The importance of data-selection criteria: meta-analysis of stream predation experiments. *Ecology* **80**, 1132–1141 (1999).
- Draper, N. R. & Smith, H. *Applied Regression Analysis* 2nd edn (Wiley and Sons, New York, 1981).
- Heyer, W. R., Donnelly, M. A., McDiarmid, R. W., Hayak, L. A. C. & Foster, M. S. (eds) *Measuring and Monitoring Biological Diversity: Standard Methods for Amphibians* (Smithsonian Institution Press, Washington, 1994).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank all the researchers who contributed amphibian population data and, in particular, J. Loman, T. Beebe and S. Zumbach for large unpublished data sets. This work was supported by Wildlife Habitat Canada, the Natural Sciences and Engineering Research Council of Canada, and the Schweizerischer Nationalfonds.

Correspondence and requests for materials should be addressed to J.E.H. (e-mail: jeffhoul@science.uottawa.ca).

The metapopulation capacity of a fragmented landscape

Ilkka Hanski & Otso Ovaskainen

Department of Ecology and Systematics, PO Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland

Ecologists and conservation biologists have used many measures of landscape structure^{1–5} to predict the population dynamic consequences of habitat loss and fragmentation^{6–8}, but these measures are not well justified by population dynamic theory. Here we introduce a new measure for highly fragmented landscapes, termed the metapopulation capacity, which is rigorously derived from metapopulation theory and can easily be applied to real networks of habitat fragments with known areas and connectivities. Technically, metapopulation capacity is the leading eigenvalue of an appropriate 'landscape' matrix. A species is predicted to persist in a landscape if the metapopulation capacity of that landscape is greater than a threshold value determined by the properties of the species. Therefore, metapopulation capacity can conveniently be used to rank different landscapes in terms of their capacity to support viable metapopulations. We present an empirical example on multiple networks occupied by an

endangered species of butterfly. Using this theory, we may also calculate how the metapopulation capacity is changed by removing habitat fragments from or adding new ones into specific spatial locations, or by changing their areas. The metapopulation capacity should find many applications in metapopulation ecology, landscape ecology and conservation biology.

The most basic aspects of metapopulation persistence in fragmented landscapes have been analysed with simple spatially implicit models⁹, akin to models of infectious disease in homogeneous populations¹⁰. In the Levins model¹¹, assuming that a fraction h of habitat patches is suitable for occupancy, the equilibrium fraction of occupied patches (out of suitable patches) is given by

$$p^* = 1 - \frac{\delta}{h} \quad (1)$$

where δ is the ratio of the extinction and colonization rate parameters: $\delta = e/c$ (refs 9 and 12–15). A well known limitation of this and other spatially implicit models is that they cannot be used to analyse explicit spatial patterns. For example, if we assume that colonization is distance-dependent, simulation studies have shown that it makes a big difference to metapopulation persistence whether habitat loss (decreasing value of h) occurs randomly or non-randomly in space^{16,17}.

A spatially realistic version of the Levins model for a finite

number of habitat patches of known areas and spatial locations can be constructed by modelling the rate of change in the probability of patch i being occupied as¹⁸:

$$\frac{dp_i(t)}{dt} = (\text{Colonization rate}_i)[1 - p_i(t)] - (\text{Extinction rate}_i)p_i(t) \quad (2)$$

The theory developed here is general and not restricted to particular functional forms of the colonization and extinction rates, but for the purpose of illustration we use here specific assumptions that are simple and biologically well justified, namely Extinction rate _{i} = e/A_i and Colonization rate _{i} = $c\sum_{j \neq i} \exp(-\alpha d_{ij})A_j p_j(t)$, where A_i is the area of patch i , d_{ij} is the distance between patches i and j , $1/\alpha$ is the average migration distance, and e and c are constants (for justification, see Methods). With these or other comparable assumptions, the most important metapopulation processes, colonization and extinction, can be related to the most important structural features of fragmented landscapes, patch areas and spatial locations^{9,19}.

We turn to matrix notation to describe the system of equations giving the rates of change for the p_i values. Using the above assumptions in equation (2), matrix $\mathbf{M}$ consists of elements $m_{ij} = \exp(-\alpha d_{ij}) A_j A_i$ for $j \neq i$ and $m_{ii} = 0$. It can be shown that an equilibrium solution with $p_i^* > 0$ for all i exists if and only if

$$\lambda_M > \delta \quad (3)$$

where λ_M is the leading eigenvalue of matrix $\mathbf{M}$. Equation (3) thus gives the condition for persistence of a species in a given landscape.

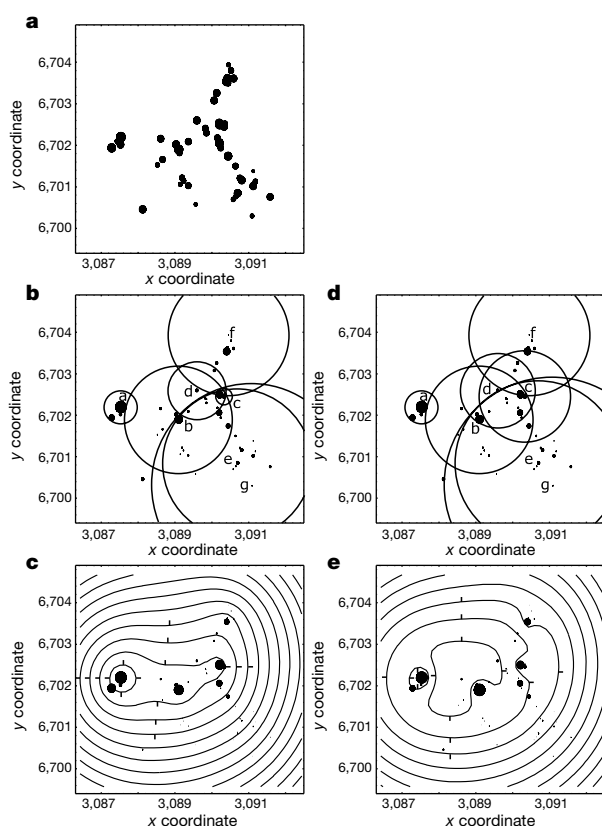


Figure 1 Change in metapopulation capacity due to removal of a patch or addition of a new patch in a particular spatial location within the existing patch network. Data based on equation (6). **a–c**, The sizes of the dots (habitat patches) are proportional to the logarithm of patch area (**a**), patch area (**b**) and the value of λ_i (**c**). The contour lines in **c** indicate the level of increase (decrease) in metapopulation capacity, corresponding to a 30% difference in patch areas, that would result from placing a new patch (removing an existing patch) in a particular location. For the explanation of the circles around patches a–g in **b** see Methods and Fig. 2. **d, e**, As in **b** and **c**, but here the model includes regional stochasticity, implemented as explained in Methods ($\gamma = 1$, $\beta = \alpha = 1$). The tick marks on the contour lines indicate the direction of the slope. x coordinate, longitude; y coordinate, latitude.

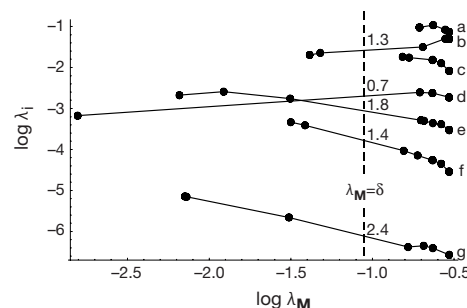


Figure 2 The contribution of patch i to λ_M , $\lambda_i \equiv \lambda_i^2 / \lambda_M$, calculated for increasingly large regions around patches a–g in Fig. 1b (see Methods). The corresponding pairs of λ_i and λ_M are shown by dots from left to right, joined by a line for each patch. The broken vertical line gives the threshold condition estimated in Fig. 3, defining the size of the circle around the patch in Fig. 2b.

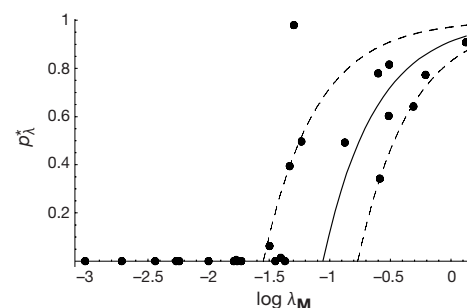


Figure 3 Plot of p_i^* against the logarithm of metapopulation capacity (λ_M) in 25 real patch networks that are potentially occupied by the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands in southwest Finland²⁰. The continuous line is based on the average of the estimated δ values for networks with $p_i^* > 0.3$. The broken lines give the minimum and maximum estimates, but omitting the two networks yielding the most extreme values.

An analogous threshold condition is well established in epidemiological theory^{10,20} for the spread of an infectious disease, although the epidemiological models are generally structured by factors such as sex or age rather than by the spatial location of populations. The threshold condition has been previously discussed in the context of metapopulation dynamics by Adler and Nüernerger²¹.

An appropriately weighted average of the p_i^* values can be closely approximated by

$$p_\lambda^* = 1 - \frac{\delta}{\lambda_M} \quad (4)$$

We note the structural identity between equations (1) and (4). In the spatially realistic model, λ_M plays exactly the same role as h , the amount of suitable habitat, plays in the simple Levins model (equation 1), in which all patches are identical and equally connected, and in which the spatial arrangement of patches therefore makes no difference. We term λ_M the metapopulation capacity of a fragmented landscape: λ_M is a measure that captures the impact of landscape structure—the amount of habitat and its spatial configuration—on metapopulation persistence. λ_M has to exceed a threshold value that is set by the properties of the species (δ) for long-term persistence (equation 3). To compute λ_M for a particular landscape, only the spatial scale of connectivity (set by the species parameter α) and the areas and the spatial locations of the habitat fragments need to be known. For a given landscape, λ_M increases with decreasing α , because small α strengthens connectivity. When α is very small, patches will contribute to λ_M in relation to their areas only. For a given species (constant α), λ_M allows a straightforward comparison of multiple fragmented landscapes, which can be conveniently ranked in terms of their capacity to support a viable metapopulation. Such comparisons are essential for management-oriented and spatially extended population viability analyses^{22,23}.

The metapopulation capacity is, to a good approximation, a sum of contributions from individual habitat fragments. The contribution of fragment i is given by $\lambda_i \equiv x_i^2 \lambda_M$, where x_i is the i th element in

the leading eigenvector of matrix $\mathbf{M}$ (see Methods). λ_i thus measures the significance of habitat fragment i to the threshold condition for metapopulation persistence. Conveniently, we can also assess how adding a new patch with area A_k to a specific location in the landscape would increase λ_M (Fig. 1; see Methods).

The above results should find useful applications in the development of algorithms for designing nature reserves^{24,25}, where the spatial dynamics of the focal species in the reserve network have typically not been considered. In the epidemiological context, λ_i gives the contribution of the given group of individuals to the spread of the disease, and could thus be used to rank, for example, vaccination scenarios. However, it is necessary to pay close attention to what λ_i really measures—the contribution of fragment i to λ_M at the threshold for deterministic persistence. In a large network with aggregated distribution of habitat patches, one particular cluster of patches will be the stronghold for the metapopulation in the entire network, and hence the properties of this cluster will largely set the threshold condition. The spatially more localized significance of individual patches can be examined by, for example, calculating the λ_M and λ_i values for smaller and larger regions around each patch (Fig. 2; see Methods). Patches that are located in the less significant patch clusters from the viewpoint of metapopulation persistence in the entire network may nonetheless have substantial significance in their own neighbourhood (for example, patches e and f in Figs 1 and 2). It is also possible to define and compute measures related to λ_M that characterize the impact of landscape structure on metapopulation invasion or on metapopulation size rather than on metapopulation persistence (O.O. & I. H., manuscript in preparation).

We now return to the nature of the weighted average of patch occupancy probabilities in equation (4). This equation is obtained when the p_i^* values are weighted by the relative contributions of the habitat fragments to metapopulation capacity, that is, $p_\lambda^* = \sum \lambda_i p_i^* / \lambda_M$. Thus p_λ^* gives the fraction of λ_M that is 'used' by the species at equilibrium. Following the general notion²⁶ that the threshold condition for persistence of a consumer is given by the amount of unused limiting resource at equilibrium, we can calculate the threshold value of λ_M for metapopulation persistence as $\delta = \lambda_M(1 - p_\lambda^*)$, which gives a heuristic derivation of equation (4).

As an example, Fig. 3 shows p_λ^* against λ_M for 25 real patch networks potentially occupied by the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands in southwest Finland. Using the formula $\delta = \lambda_M(1 - p_\lambda^*)$, the threshold value was estimated for networks with $p_\lambda^* > 0.3$ (see Methods). The average value of these relatively independent estimates predicts well the absence or near absence of the species in the remaining networks (continuous line in Fig. 3).

Our model can also be used to examine the consequences of habitat loss on metapopulation persistence. Let p_A^* be the average patch occupancy probability weighted by patch area. p_A^* gives a reasonably good approximation of p_λ^* , though it tends to give an overestimate (O.O. & I. H., manuscript in preparation). The exact value of p_A^* can be calculated by first solving the values of p_i^* by iteration²⁷. If h_A is the fraction of the pooled habitat area that remains suitable for occupancy, the fraction of unused habitat (out of the original amount of habitat) at equilibrium is given by

$$\text{Amount of empty habitat} = h_A(1 - p_A^*) \approx \frac{h_A \delta}{\lambda_M} \quad (5)$$

(assuming that $p_A^* \approx p_\lambda^*$). In the Levins model the amount of empty habitat remains constant for $h > \delta$ (see equation (1) and refs 9, 12 and 14). This is approximately so in the present model if habitat loss is random, in which case λ_M decreases roughly in proportion to the decrease in the amount of suitable habitat, h_A (Fig. 4b). In contrast, if habitat is lost in large blocks, λ_M decreases initially less than in proportion to the decrease in h_A , as metapopulation dynamics in the remaining habitat are affected relatively little (Fig. 4d). Therefore, such non-random loss of habitat is less detrimental to metapopulation

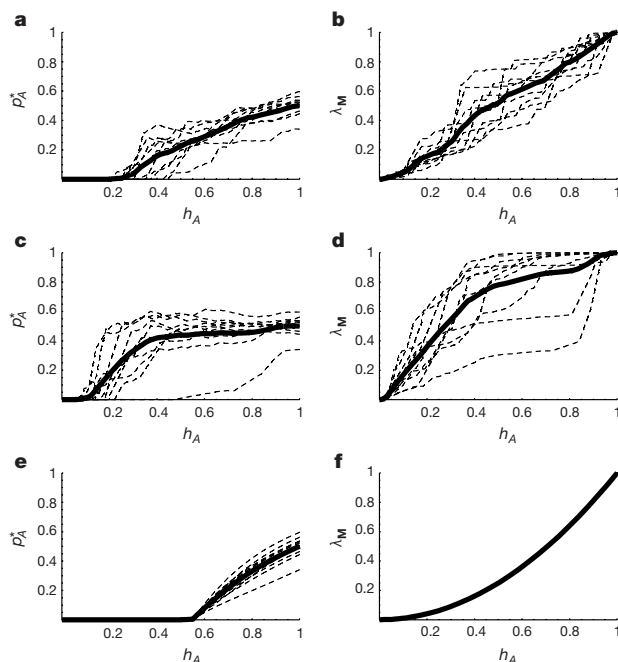


Figure 4 Consequences of habitat loss on patch occupancy and the threshold value for metapopulation persistence. Three scenarios are shown: **a, b**, random loss of habitat patches; **c, d**, systematic (non-random) loss of patches; **e, f**, loss of area from each existing patch. For each scenario the fraction of occupied habitat (p_A^*) is plotted against the fraction of suitable habitat (h_A) (**a, c, e**) and the relationship between λ_M and h_A (see Methods) is shown (**b, d, f**). Solid lines, average.

persistence than random loss of habitat (compare Figs. 4a and c; this conclusion must naturally be applied cautiously to real landscapes and populations, because they may be affected by processes not included in the model). The present model can be used to investigate the consequences of any pattern of habitat loss, for example, decreasing areas of the existing fragments, which is often even more harmful to metapopulation persistence (Fig. 4e) than random loss of entire patches (Fig. 4a).

The examples presented here illustrate how metapopulation capacity can be used to examine both the local and the global aspects of metapopulation persistence in real fragmented landscapes. We expect that metapopulation capacity will find many constructive applications in metapopulation ecology, landscape ecology and conservation biology. □

Methods

The model structure

We have used here a specific spatially realistic version of the Levins model to illustrate the concept of metapopulation capacity. The numerical results are model-specific, because λ_M integrates the effects of landscape structure on metapopulation processes as specified by a dynamic model, but the theory is general and results can be readily obtained for a range of continuous-time and discrete-time models with different structural assumptions. The general mathematically rigorous theory will be presented elsewhere (O.O. & I. H., manuscript in preparation).

The present structural model assumptions are justified as follows. The extinction rate is a function of the inverse of patch area, e/A_p , because large patches tend to have large expected population sizes and because extinction risk scales roughly as the inverse of the expected population size in taxa affected by moderate environmental stochasticity^{9,28,29}. The colonization rate is given as a sum of contributions from the existing populations, $c\sum_{j \neq i} \exp(-\alpha d_{ij}) A_j p_j(t)$, because immigration to patch i is expected to increase with the number of neighbouring populations, with their sizes as reflected by the respective patch areas, and with their decreasing distances to the focal patch and increasing incidences of occupancy.

Variation in patch quality and regional stochasticity

Instead of using patch area as a surrogate of expected population size, patch areas may be corrected, given sufficient information, for spatially varying habitat quality. Also, a simple but effective way of incorporating the consequences of regional stochasticity (spatially correlated environmental stochasticity) into the model is to replace the term $p_j(t)$ in the expression for colonization rate of patch i with a term such as $(1 - \gamma \exp(-\beta d_{ij})) p_j(t)$. This term takes into account the fact that if extinctions are caused by regional stochasticity, the probabilities of patches i and j being empty become increasingly positively correlated. In addition, a negative correlation increasing λ_M may emerge if there exists an interaction between habitat quality and regional stochasticity (different patches are 'good' and 'bad' in different years).

Relative patch values

The contribution of patch i to λ_M can be closely approximated by $\lambda_i \equiv x_i^2 \lambda_M$ (O.O. & I. H., manuscript in preparation) where x_i is the i th element in the leading eigenvector of matrix $\mathbf{M}$ (the values of x_i^2 are scaled to sum up to unity). The addition to λ_M that would be obtained by adding a new patch with area A_k to a specific location in the landscape can be calculated as

$$\lambda_k = \frac{A_k^2}{\lambda_M} \left(\sum_{j \neq k} e^{-\alpha d_{jk}} A_j x_j \right)^2 \quad (6)$$

where d_{jk} is the distance from the existing patch j to the hypothetical patch k . Note that λ_k is defined as a product of contributions from patch area and spatial location, hence the advantage of having a new patch in a particular spatial location in the landscape can be evaluated independently of its area (Fig. 1: the network used in this example is one of the networks occupied by the Glanville fritillary butterfly in the Åland Islands in southwest Finland³⁰).

The local value of patches labelled a–g in Figs. 1b and d was calculated for circular areas centred on the respective patch and with a radius of 0.5, 1.0, 1.5, ... ($\alpha = 1$). For each circle, the values of λ_i and λ_M were calculated and displayed in Fig. 2. At the largest scale, when the circle around each patch includes all the other patches, the values converge to the global values shown in Fig. 1c. The intersection point of each patch-specific line joining the points (λ_i, λ_M) with the threshold condition (broken line in Fig. 2) gives the radius of the circle around patch i within which the threshold condition for metapopulation persistence is satisfied; in the case of patches a and c the threshold condition is satisfied with the smallest radius, 0.5. Figure 1b shows these areas for patches a–g. In Fig. 1d, which shows the result with regional stochasticity (see above in Methods), the radius of the circle corresponds to the same ratio δ/λ_M as in Fig. 1b.

Parameter estimation

For a metapopulation at stochastic steady state, the model can be parameterized using the formula $\delta = \lambda_M(1 - p_\lambda^*)$. In the example in Fig. 3, the value of p_λ^* was calculated based on

patch areas, spatial locations and the occurrence of the butterfly in the patches in 1993, the last providing (rough) empirical estimates of the p_i^* values. Only networks in the western Åland were included (network mid-point west of the longitude 310400 in the Finnish Uniform Coordinate System), as our previous analyses have indicated that many metapopulations in the eastern Åland were either severely out of the steady state or there are some environmental differences influencing patch occupancy³⁰. For calculating λ_M , the value of $\alpha = 1$ was used, as estimated in mark–release–recapture studies³⁰.

Habitat loss

The results in Fig. 4 were calculated for hypothetical landscapes with 100 randomly located patches within an area of 5 by 5 units. The systematic loss of patches was obtained by reducing the area of the square from the margins and eliminating the patches that were located outside the reduced area. Patch areas are log-normally distributed. Model parameters have the values $\alpha = 1$ and $\delta = 0.3$. Results were calculated for 10 replicates, with the thick line showing the average. Patch areas in the replicates were scaled to give $\lambda_M = 1$.

Received 10 November 1999; accepted 14 February 2000.

- Turner, M. G., O'Neill, R. V., Gardner, R. H. & Milne, B. T. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecol.* **3**, 153–162 (1989).
- Turner, M. G. & Gardner, R. H. (eds) *Quantitative Methods in Landscape Ecology* (Springer, New York, 1991).
- Gardner, R. H., O'Neill, R. V. & Turner, M. G. in *Humans as Components of Ecosystems: Subtle Human Effects and the Ecology of Population Areas* (eds Pickett, S. T. A. & McDonnell, M. J.) 208–226 (Springer, New York, 1993).
- Lavorel, S., Gardner, R. H. & O'Neill, R. V. Analysis of patterns in hierarchically structured landscapes. *Oikos* **67**, 521–528 (1993).
- Gustafsson, E. J. Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems* **1**, 143–156 (1998).
- Bascompte, J. & Solé, R. V. Spatially induced bifurcations in single-species population dynamics. *J. Anim. Ecol.* **63**, 256–264 (1994).
- Gustafsson, E. J. & Gardner, R. H. The effect of landscape heterogeneity on the probability of patch colonization. *Ecology* **77**, 94–107 (1996).
- With, K. A., Gardner, R. H. & Turner, M. G. Landscape connectivity and population distributions in heterogeneous landscapes. *Oikos* **78**, 151–169 (1997).
- Hanski, I. *Metapopulation Ecology* (Oxford Univ. Press, Oxford, 1999).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
- Levins, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15**, 237–240 (1969).
- Lande, R. Extinction thresholds in demographic models of territorial populations. *Am. Nat.* **130**, 624–635 (1987).
- May, R. M. The role of ecological theory in planning reintroduction of endangered species. *Symp. Zool. Soc. London* **62**, 145–163 (1991).
- Nee, S. & May, R. M. Dynamics of metapopulations: habitat destruction and competitive coexistence. *J. Anim. Ecol.* **61**, 37–40 (1992).
- Lawton, J. H., Nee, S., Letcher, A. J. & Harvey, P. in *Large-scale Ecology and Conservation Biology* (eds Edwards, P. J., May, R. M. & Webb, N. R.) 41–58 (Blackwell Scientific, Oxford, 1994).
- Hill, M. F. & Caswell, H. Habitat fragmentation and extinction thresholds on fractal landscapes. *Ecol. Lett.* **2**, 121–127 (1999).
- With, K. A. & King, A. W. Extinction thresholds for species in fractal landscapes. *Conserv. Biol.* **13**, 314–326 (1999).
- Hanski, I. & Gyllenberg, M. Uniting two general patterns in the distribution of species. *Science* **275**, 397–400 (1997).
- Hanski, I. A practical model of metapopulation dynamics. *J. Anim. Ecol.* **63**, 151–162 (1994).
- Grenfell, B. T. & Dobson, A. P. (eds) *Ecology of Infectious Diseases in Natural Populations* (Cambridge Univ. Press, Cambridge, 1995).
- Adler, F. R. & Nuerberger, B. Persistence in patchy irregular landscapes. *Theor. Pop. Biol.* **45**, 41–75 (1994).
- Lindemayer, D. B. & Possingham, H. P. *The Risk Of Extinction: Ranking Management Options For Leadbeater's Possum Using Population Viability Analysis* (Centre for Resource and Environmental Studies, Canberra, 1994).
- Hanski, I. in *The Ecological Basis of Conservation* (eds Pickett, S. T. A., Ostfeld, R. S., Shachak, M. & Likens, G. E.) 217–227 (Chapman & Hall, New York, 1997).
- Freitag, S. & van Jaarsveld, A. S. Sensitivity of selection procedures for priority conservation areas to survey extent, survey intensity and taxonomic knowledge. *Proc. R. Soc. Lond. B* **265**, 1475–1482 (1998).
- Pressey, R. L., Possingham, H. P. & Margules, C. R. Optimality in reserve selection algorithms: when does it matter and how much? *Biol. Conserv.* **76**, 259–267 (1996).
- Nee, S. How populations persist. *Nature* **367**, 123–124 (1994).
- Hanski, I. Metapopulation dynamics. *Nature* **396**, 41–49 (1998).
- Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **142**, 911–927 (1993).
- Foley, P. Predicting extinction times from environmental stochasticity and carrying capacity. *Conserv. Biol.* **8**, 124–137 (1994).
- Hanski, I., Moilanen, A., Pakkala, T. & Kuussaari, M. The quantitative incidence function model and persistence of an endangered butterfly metapopulation. *Conserv. Biol.* **10**, 578–590 (1996).

Acknowledgements

We thank N. Behera, K. Frank, B. Grenfell, M. Keeling, A. Moilanen, B. O'Hara, T. Roslin, I. Saccheri and C. Thomas for comments on the manuscript.

Correspondence and requests for materials should be addressed to I. H. (e-mail: ilkka.hanski@helsinki.fi).

Asymmetric redirection of flow through the heart

Philip J. Kilner*, Guang-Zhong Yang*†, A. John Wilkes‡, Raad H. Mohiaddin*, David N. Firmin* & Magdi H. Yacoub§

* Cardiovascular Magnetic Resonance Unit, † Visual Information Processing Group of Department of Computing and § Department of Cardiothoracic Surgery, Royal Brompton Hospital site of Imperial College of Science, Medicine and Technology, Sydney Street, London SW3 6NP, UK

‡ Flow Design Research Group, Emerson College, Forest Row, East Sussex RH18 5JX, UK

Through cardiac looping during embryonic development¹, paths of flow through the mature heart have direction changes and asymmetries whose topology and functional significance remain relatively unexplored. Here we show, using magnetic resonance velocity mapping^{2–5}, the asymmetric redirection of streaming blood in atrial and ventricular cavities of the adult human heart, with sinuous, chirally asymmetric paths of flow through the whole. On the basis of mapped flow fields and drawings that illustrate spatial relations between flow paths, we propose that asymmetries and curvatures of the looped heart have potential fluidic and dynamic advantages. Patterns of atrial filling seem to be asymmetric in a manner that allows the momentum of inflowing streams to be redirected towards atrio-ventricular valves, and the change in direction at ventricular level is such that recoil away from ejected blood is in a direction that can enhance rather than inhibit ventriculo-atrial coupling⁶. Chiral asymmetry might help to minimize dissipative interaction between entering, recirculating and outflowing streams⁷. These factors might combine to allow a reciprocating, sling-like, 'morphodynamic' mode of action to come into effect when heart rate and output increase during exercise⁶.

Although flow in the human heart, particularly the left ventricle, has been investigated by physical and computational modelling^{8–10}, and by invasive and non-invasive imaging^{11,12}, there is still only limited understanding of changes in direction and asymmetries of flow through heart cavities, or of their potential significance in relation to function. Vertebrates show cephalo-caudal and dorso-ventral asymmetries of body plan with respect to which lateral asymmetries of the heart and other internal organs occur^{13,14}. Cardiac looping involves the ventral displacement of a ventricular relative to an atrial 'loop' in a ventrally located, cephalically directed flow path. Lateral displacements of parts of this sinuous path result in chiral, quasi-helical asymmetries. Plurality of venous inflows and septation of systemic from pulmonary flow paths complicate the picture further, but for the purposes of this paper, attention is focused on asymmetries and changes in direction of flow in heart cavities, subordinate to asymmetries of the heart as a whole.

We used magnetic resonance phase-velocity mapping^{3,4} to investigate patterns of flow in human heart cavities, aiming at visualizing cavity filling patterns and elucidating major changes in direction of blood. The technique, besides using transiently applied magnetic gradients to locate signal from nuclei in three-dimensional space, uses gradient applications to encode velocity in the phase of the radio signal received back from energized nuclei. The nuclei in question are those of hydrogen in water, so no contrast agent or marker is needed for this non-invasive flow-imaging technique. The flow images, reconstructed from the signal received over many heart cycles, record large-scale flow features that recur repeatedly through successive beats. Figure 1 shows systolic and diastolic streamline maps⁵ computed from mid-systolic and early diastolic frames of 16-phase cine magnetic resonance velocity acquisitions⁴ in a healthy 34-year-old man in planes located through the cavities of the right

atrium, the left atrium and the left ventricle. Corresponding animated streamline maps were also made (see Supplementary Information). For assessment of consistency of major flow features between subjects, velocity maps were acquired in 21 additional healthy volunteers (see Tables 1 and 2 of Supplementary Information). Topologies and orientations of large-scale intra-cavity flow features

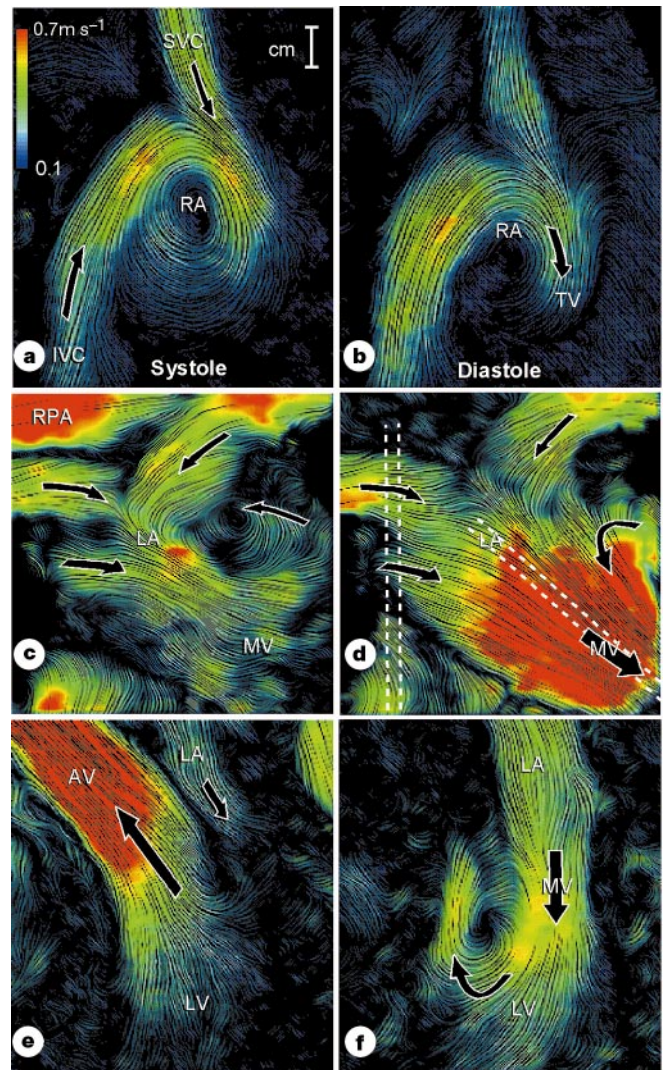


Figure 1 Asymmetric intracardiac flow patterns. **a**, In the right atrium in ventricular systole, viewed in a sagittal plane from the subject's right side, blood entering from superior and inferior caval veins contributes to the forward rotation of blood in the expanding chamber. Coloured streamlines computed from magnetic resonance velocity acquisitions show local speed, as indicated by the colour scale. **b**, In early ventricular diastole, further inflow of blood is again redirected forwards and down the front of the right atrium, and away from the viewer (out of plane) through the open tricuspid valve. **c**, In the left atrium in ventricular systole, blood enters from the upper and lower pulmonary veins on each side, indicated by arrows in this coronal plane viewed from the front (lower veins lie out of plane posteriorly). Streamlines are redirected asymmetrically round towards the mitral valve, which is closed, but pulled by contraction of the left ventricle. **d**, In early ventricular diastole, there is further inflow from veins while blood passes through the open mitral valve to the left ventricle. Vertical and oblique dotted lines indicate planes orthogonal to this panel represented by panels **a**, **b**, **e** and **f**. **e**, In the left ventricle in systole, streamlines pass from the left ventricle through the aortic valve, viewed here in an oblique long-axis plane from above left, the front of the subject being located to the left of the image. **f**, In early diastole, streamlines pass from the left atrium through the open mitral valve to the left ventricle, with asymmetric recirculation (curved arrow) round the anterior leaflet of the mitral valve. In **e** and **f** only, the colour scale is modified to reach red at 1 m s^{-1} . AV, aortic valve; IVC, inferior vena cava; LA, left atrium; LV, left ventricle; MV, mitral valve; RA, right atrium; RPA, right pulmonary artery; SVC, superior vena cava; TV, tricuspid valve.

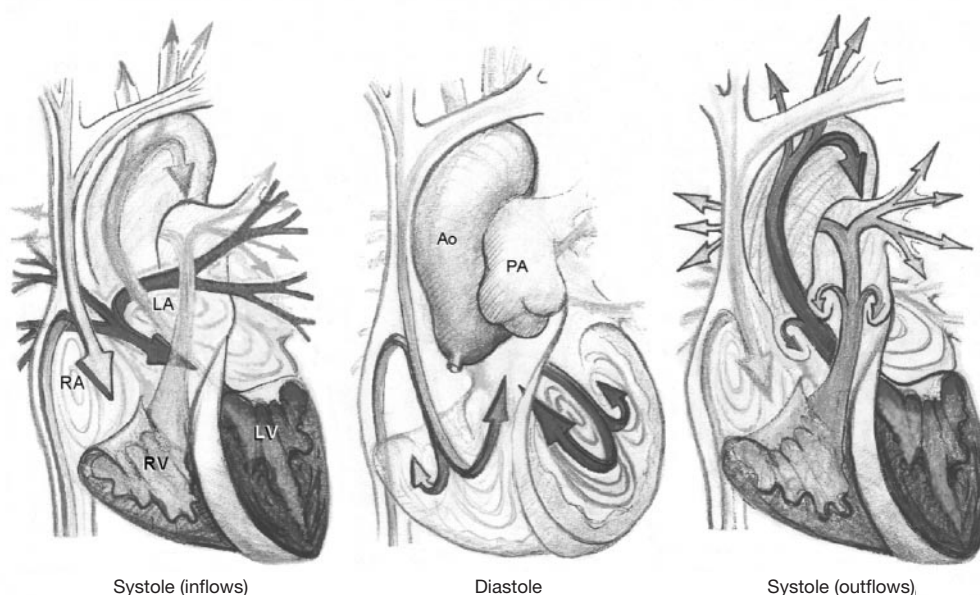


Figure 2 Changes in direction of blood through the heart. Main streams in systolic and early diastolic phases have been drawn as viewed from the front. Systole is shown twice so that the passage of blood can be followed from atrial inflows, to ventricles, to arterial

outflows. Ao, aorta; LA, left atrium; LV, left ventricle; PA, pulmonary artery; RA, right atrium; RV, right ventricle.

were broadly similar between volunteers studied (see Supplementary Information).

In the right atrium (Fig. 1a, b), streams from superior and inferior caval veins did not collide head on but turned forward, contributing to a forward rotation of blood, turning clockwise as viewed from the subject's right side. This filling pattern was associated with movement of the anterior part of the right atrial blood volume towards the inlet of the tricuspid valve. Inflows to both right and left atrial cavities showed peaks in two phases, coinciding with ventricular systole (Fig. 1a, c) and early ventricular diastole (Fig. 1b, d).

In the left atrium (Fig. 1c, d), imaging in a coronal slice showed that inflows from pulmonary veins, those on the left located slightly higher than those on the right, contributed to net rotational momentum—anticlockwise as viewed from the front. The asymmetric filling pattern was associated with leftwards and downwards movement of the inferior half of the contained blood volume, across towards the mitral valve, in systole as well as in diastole.

In the left ventricle, imaging in an oblique long-axis plane showed ejection of blood to the aorta in systole (Fig. 1e) and refilling from the left atrium in early diastole (Fig. 1f). There was a second, late diastolic phase of ventricular filling in the resting state⁶, coinciding with atrial systole, not shown here. Inflow through the open mitral valve gave rise to recirculating flows beneath the valve leaflets, the dominant direction being under the free edge of the anterior mitral leaflet. Part of the blood volume was thus redirected towards the outflow tract (Fig. 1f). Transient recirculation was also seen beneath the posterior mitral valve leaflet. An almost complete change of direction between left ventricular inflow and outflow is shown by comparison of Fig. 1f and Fig. 1e.

In the right ventricle, inflow through the tricuspid valve also gave rise to recirculating flows beneath the leaflets, the dominant direction being towards the outflow region. The change in direction between inflow and outflow was less acute than that in the left ventricle. However, the right ventricle has a more marked chiral asymmetry than the left, its volume being indented by curvature of the septum, making the depiction of right ventricular flow in a single plane unsatisfactory.

Figure 2 gives an overview of systolic and early diastolic flows of the whole heart, gained from velocity map data acquired in five

contiguous coronal planes covering the volume of the heart in the same subject. Flow paths were traced by hand on a transparent sheet taped to the computer screen as velocity maps were displayed one by one, working from the back of the heart to the front. The band-like depiction of principal flow paths allows overall spatial relationships and chiral asymmetries, not apparent in two-dimensional streamline maps, to be visualized. The drawing also permits recognition of potential continuity of momentum between chambers and between phases, which is relevant when, on exercise, systole and diastole alternate in rapid succession.

In a previous echocardiographic study⁶, we documented a transition from biphasic left ventricular filling at rest, to rapid, monophasic filling during strenuous exercise. M-mode and Doppler echocardiographic traces showed evidence of enhanced, rapid reciprocation of atrial and ventricular action on exertion. During strenuous exercise, atrial systole came to coincide with early diastolic filling in a single short diastolic period of rapid ventricular filling. As the heart rate more than doubled, atrial systole (ventricular diastole) and ventricular systole (atrial diastole) alternated rapidly, peak velocities of ventricular inflow and outflow rising to about twice those of the resting state. These changes are compatible with the marked enhancement, during exercise, of exchanges of force associated with changes in momentum through curvatures of the heart.

Figure 3 shows drawings illustrating theoretical fluidic and dynamic consequences of looped as opposed to linear arrangements of deliberately simplified two-chamber heart models without chiral asymmetry. Whereas vertebrates have looped arrangements of juxtaposed atrio-ventricular cavities, a nearly linear arrangement occurs in dorsally located hearts of snails such as *Helix pomatia*—animals not noted for dynamic vigour. The depiction of intracavity recirculation patterns in Fig. 3 is based on the visualization of flow in symmetrical and asymmetric physical flow models (see animated images in Supplementary Information). The term 'fluidic' refers to the participation of the fluid's own dynamics in the control of direction and timing of flow through the system.

On the basis of mapped intracardiac flow patterns, we propose that curvatures and asymmetries of the heart confer potential functional advantages that could gain importance as flow velocities, heart rate and rates of change of momentum increase with exertion: eccentric alignments of venous inflows with respect to atrial cavities

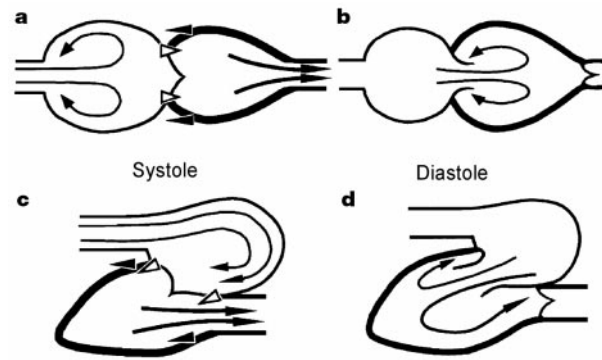


Figure 3 Comparison of linear and looped atrio-ventricular arrangements. **a**, In the linear arrangement in systole, contraction of the ventricle (thick boundary) pulls on the atrium (white arrowheads), contributing to atrial expansion. Inflow gives rise to recirculation bilaterally (thin arrows), a relatively unstable pattern of flow that redirects blood inappropriately for subsequent ventricular filling. Any recoil of the ventricle (black arrowheads) away from blood accelerated to the outflow (thick arrows) would push back against the atrium, counteracting atrial expansion. **b**, In diastole, recirculation in the expanding ventricle redirects fluid away from the outflow tract. **c**, In a sinuously looped arrangement in systole, ventricular contraction also expands the adjacent atrium (white

arrowheads). Atrial filling is now asymmetric and more stable, streamlines being accommodated by wall curvatures in a way that redirects momentum towards the atrio-ventricular valve. Recoil of the ventricle away from ejected blood (black arrowheads) now adds to the pull on the atrio-ventricular junction, enhancing rather than suppressing atrial expansion. **d**, In diastole, redirected intra-atrial momentum can contribute to ventricular filling, which occurs with asymmetric recirculation, redirecting flow preferentially towards the outflow tract. Looped curvature thus allows the sinuous redirection of momentum and dynamically enhanced reciprocation between atrial and ventricular function.

predispose to asymmetry of intra-atrial flow, redirecting inflow towards rather than away from atrio-ventricular valves (Figs 1–3). Relatively coherent swirling of blood, although potentially associated with higher wall shear stresses, might avoid excessive dissipation of energy by limiting flow separation and instability (see animated images in Supplementary Information). Chiral or non-planar asymmetry (Fig. 2) might also avoid instabilities by allowing entering, recirculating and outflowing streams to pass one another in three-dimensional space without collision⁷. At ventricular level, change in direction is such that recoil away from ejected blood is in a direction that can enhance rather than inhibit ventriculo-atrial coupling (Figs 1–3). Combining these effects, we propose that, during exercise, the looped heart is able to function ‘morphodynamically’, redirecting and slinging blood through its sinuous curvatures with minimal dissipation of energy and with dynamically enhanced reciprocation of atrial and ventricular function. Transition to this mode, in which forces associated with changes in momentum gain functional importance, can be likened to changes of whole-body mechanics when the actions of walking speed up to those of running. Dynamic exchanges and efficiency of the heart in this state could involve coupled interactions between contractility, elasticity and change in momentum, at atrial, ventricular and vascular levels. In health, auto-adjustments of contractility^{15,16}, compliance¹⁷ and structure^{18,19} might contribute to interrelations appropriate to morphodynamic heart action.

This interpretation has potential medical relevance where pathologies or interventions entail altered relations between form, flow, mobility and timing of the heart, especially in individuals who wish to maintain physically active lifestyles. In relation to heart surgery, the interpretation supports the proposition that spatial relations and mobility of cardiovascular tissues should be conserved or reinstated, if technically possible²⁰. The relative stability of flow through the heart might have bearing on the avoidance of thrombosis, and, by affecting flow arriving at arterial branches, possibly on the avoidance of atherogenicity²¹. Our proposed functional interpretation is of interest in relation to research into the genetic¹⁴ and phylogenetic origins of cardiac looping. Looped curvature of a ventrally located heart seems to be a feature confined to vertebrates, having apparently been retained and elaborated as classes evolved²². The ability of the looped heart to deliver enhanced output during strenuous exertion might have been a factor permitting the evolution of large, complex, dynamically active species characteristic of the vertebrate line. □

Received 6 December 1999; accepted 9 February 2000.

1. Icardo, J. M. Development biology of the vertebrate heart. *J. Exp. Zool.* **275**, 144–161 (1996).
2. Firmin, D. N. *et al.* In vivo validation of MR velocity imaging. *J. Comput. Assist. Tomogr.* **11**, 751–756 (1987).
3. Firmin, D. N., Nayler, G. L., Kilner, P. J. & Longmore, D. B. The application of phase shifts in NMR for flow measurement. *Magn. Reson. Med.* **14**, 230–241 (1990).
4. Kilner, P. J., Yang, G. Z., Mohiaddin, R. H., Firmin, D. N. & Longmore, D. B. Helical and retrograde secondary flow patterns in the aortic arch studied by three-directional magnetic resonance velocity mapping. *Circulation* **88**, 2235–2247 (1993).
5. Yang, G. Z., Kilner, P. J., Mohiaddin, R. H. & Firmin, D. N. Transient streamlines: texture synthesis for in vivo flow visualisation. *Int. J. Cardiac Imag.* (in the press).
6. Kilner, P. J., Henein, M. Y. & Gibson, D. G. Our tortuous heart in dynamic mode—an echocardiographic study of mitral flow and movement in exercising subjects. *Heart Vessels* **12**, 103–110 (1997).
7. Caro, C. G. *et al.* Non-planar curvature and branching of arteries and non-planar-type flow. *Proc. R. Soc. Lond. A* **452**, 185–197 (1996).
8. Bellhouse, B. J. & Bellhouse, F. H. Fluid mechanics of the mitral valve. *Nature* **224**, 615–616 (1969).
9. Peskin, C. S. & McQueen, D. M. in *Case Studies in Mathematical Modelling—Ecology, Physiology, and Cell Biology* (eds Othmer, H. G., Adler, F. R., Lewis, M. A. & Dallon, J. C.) 309–337 (Prentice-Hall, Eaglewood Cliffs, NJ, 1996).
10. Taylor, T. W. & Yamaguchi, T. Flow patterns in three-dimensional left ventricular systolic and diastolic flows determined from computational fluid dynamics. *Biorheology* **32**, 61–71 (1995).
11. Taylor, D. E. & Wade, J. D. Pattern of blood flow within the heart: a stable system. *Cardiovasc. Res.* **7**, 14–21 (1973).
12. Kim, W. Y. *et al.* Left ventricular blood flow patterns in normal subjects: a quantitative analysis by three-dimensional magnetic resonance velocity mapping. *J. Am. Coll. Cardiol.* **26**, 224–238 (1995).
13. Ryan, A. K. *et al.* Pitx2 determines left–right asymmetry of internal organs in vertebrates. *Nature* **394**, 545–551 (1998).
14. Logan, M., Pagan-Westphal, S. M., Smith, D. M., Paganessi, L. & Tabin, C. J. The transcription factor Pitx2 mediates situs-specific morphogenesis in response to left–right asymmetric signals. *Cell* **94**, 307–317 (1998).
15. Noble, M. I. The Frank–Starling curve. *Clin. Sci. Mol. Med.* **54**, 1–7 (1978).
16. Landesberg, A. Molecular control of myocardial mechanics and energetics: the chemo-mechanical conversion. *Adv. Exp. Med. Biol.* **430**, 75–87 (1997).
17. Liang, Y. L. *et al.* Effects of heart rate on arterial compliance in men. *Clin. Exp. Pharmacol. Physiol.* **26**, 342–346 (1999).
18. Di Bello, V. *et al.* Left ventricular function during exercise in athletes and in sedentary men. *Med. Sci. Sports Exercise* **28**, 190–196 (1996).
19. Shapiro, L. M. The morphologic consequences of systemic training. *Cardiol. Clinics* **15**, 373–379 (1997).
20. Yacoub, M. H., Kilner, P. J., Birks, E. J. & Misfeld, M. The aortic outflow and root—a tale of dynamism and crosstalk. *Ann. Thorac. Surg.* **68**, 37–43 (1999).
21. Friedman, M. H., Brinkman, A. M., Qin, J. J. & Seed, W. A. Relation between coronary artery geometry and the distribution of early sudanophilic lesions. *Atherosclerosis* **98**, 193–199 (1993).
22. Kilner, P. J. *Morphodynamics of Flow Through the Heart*. Thesis, Univ. London (1998).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by grants from the British Heart Foundation and CODA, and by a HEFCE JREI award.

Correspondence and requests for materials should be addressed to P.J.K. (e-mail: p.kilner@rbh.nthames.nhs.uk).

Interruption of a basal ganglia–forebrain circuit prevents plasticity of learned vocalizations

Michael S. Brainard & Allison J. Doupe

Keck Center for Integrative Neuroscience, Departments of Physiology and Psychiatry, University of California San Francisco, San Francisco, California 94143-0444, USA

Birdsong, like speech, is a learned vocal behaviour that relies greatly on hearing; in both songbirds¹ and humans² the removal of auditory feedback by deafening leads to a gradual deterioration of adult vocal production. Here we investigate the neural mechanisms that contribute to the processing of auditory feedback during the maintenance of song in adult zebra finches. We show that the deleterious effects on song production that normally follow deafening can be prevented by a second insult to the nervous system—the lesion of a basal ganglia–forebrain circuit. The results suggest that the removal of auditory feedback leads to the generation of an instructive signal that actively drives non-adaptive changes in song; they also suggest that this instructive signal is generated within (or conveyed through) the basal ganglia–forebrain pathway. Our findings provide evidence that cortical–basal ganglia circuits may participate in the evaluation of sensory feedback during calibration of motor performance, and demonstrate that damage to such circuits can have little effect on previously learned behaviour while conspicuously disrupting the capacity to adaptively modify that behaviour.

The learning of birdsong, an intricate, sequenced motor skill, occurs in a two-stage process^{3,4}. First, in a sensory learning period, young birds listen to and memorize the song of an adult tutor. Then, in a sensorimotor learning period, they gradually match their own developing vocalizations to the tutor song memory. This process bears striking similarities to certain aspects of human vocal learning⁴. In particular, both song and speech are strongly dependent on hearing; neither develops normally in deaf individuals, and, in adults, hearing loss or experimental alteration of auditory feedback can lead to dramatic changes in vocal production^{1,2,5,6}. Birdsong thus provides a model system for investigating the neural processing whereby auditory feedback is used to calibrate and maintain vocal production, as well as the more general question of how performance-based feedback is used to shape complex motor output. Here we show that a basal ganglia–forebrain circuit that is not required for normal song production in adult zebra finches is nevertheless critical for plasticity in song following interruption of auditory feedback.

The rationale for our experiment is illustrated by tracing the hypothetical flow of auditory feedback through the brain as birds learn to produce and stabilize their own song (Fig. 1a–d). During normal sensorimotor learning, young birds use auditory feedback to match their own vocalizations to the previously memorized tutor song⁷. To carry out this matching process, the nervous system must first encode auditory feedback from the bird's own song, and then compare that feedback with the stored song target (tutor song). This comparison is presumed to generate an instructive signal that gradually drives adaptive changes in the motor pathway for song (Fig. 1a). At the end of sensorimotor learning, the bird's song resembles that of the tutor and in many species is retained relatively unchanged throughout adulthood.

The unchanging nature of adult birdsong might reflect a stabilization of synapses in the motor pathway for song production, such that they become independent of auditory experience once sensorimotor learning is complete. Recent experiments have shown,

however, that alteration of auditory feedback by deafening or by the playback of masking sounds can lead to gradual deterioration of the songs of adult zebra finches^{1,5}. This suggests the alternative hypothesis that the neural mechanisms involved in the evaluation of auditory feedback remain active in adult birds, but that normally there is no impetus for change in song because it matches the stored song target (Fig. 1b). According to this model, the changes in song following disruption of auditory feedback reflect an active process, in which the mismatch between feedback and the stored song target generates a large but aberrant instructive signal, which drives non-adaptive changes in song (Fig. 1c).

This interpretation raises the question of where in the brain auditory feedback of the bird's song is evaluated. One candidate is

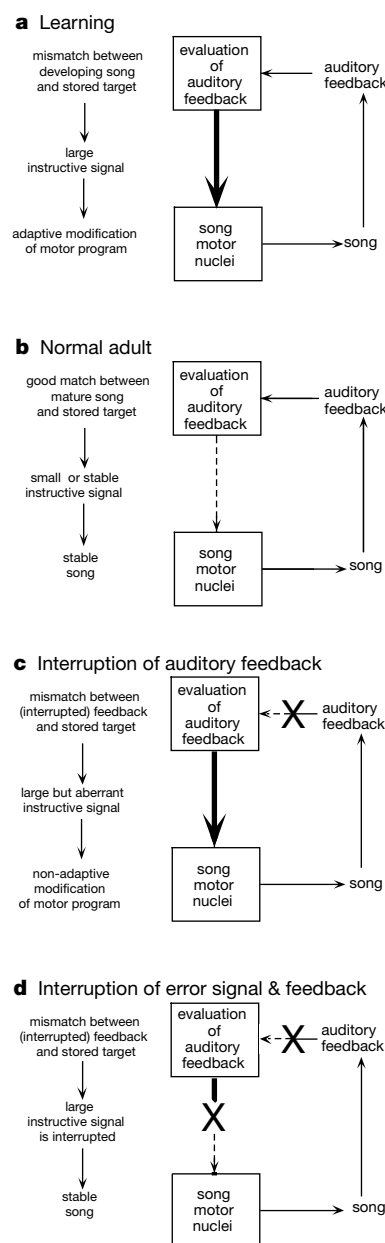


Figure 1 Hypothetical flow of auditory feedback through the song system. **a**, During sensorimotor learning, auditory feedback from song is compared with a previously memorized song target. Differences between the bird's own song and the target song result in an instructive signal that drives adaptive modification of song. **b**, After song learning, the bird's own song closely approximates the target song. Consequently, there is little drive for further changes. **c**, Removal of feedback leads to generation of an aberrant instructive signal and non-adaptive changes to song. **d**, Interruption of the instructive signal removes impetus for changes to song, even when feedback is altered.

the anterior forebrain pathway (AFP; Fig. 2a), a basal ganglia–forebrain circuit⁸ belonging to a system of brain areas devoted to vocal learning and production⁹. Lesions of this pathway in juvenile birds prevent normal song learning^{10–12}. Moreover, the AFP is well situated to guide changes in the connectivity of the motor pathway for song; the pathway originates in the song motor control nucleus HVC and passes through the basal ganglia (Area X)⁸, the thalamus (DLM), and a dorsal forebrain nucleus analogous to frontal cortex (LMAN: lateral magnocellular nucleus of the anterior neostriatum), before returning to the motor pathway for song at the level of the robust nucleus of the archistriatum (RA). Finally, neurons in the AFP respond well to the sound of the bird's own song or the tutor song, but poorly if at all to most other sounds, including the songs of other individuals of the same species^{13–15}. This auditory selectivity potentially endows the AFP with the capacity to 'listen to' and evaluate the bird's own song relative to the tutor song, and provide instructive feedback to the motor pathway via its output to RA.

Although lesions of the AFP in young birds disrupt song learning, lesions in adults have little apparent effect on song production^{9–12,16}. This lack of effect of adult lesions contrasts with the gradual deterioration of song following deafening in adult birds¹, and has led to the suggestion that the AFP cannot participate in the processing of auditory feedback during the maintenance of adult song. However, although these results indicate that lesioning the AFP is not equivalent to removing auditory feedback as it enters the

nervous system (Fig. 1c), they do not rule out the possibility that the AFP is involved in the subsequent evaluation of auditory feedback. If lesions of this pathway interrupt an instructive signal based on an evaluation of auditory feedback, they might not have an effect on normal adult song, which is well matched to its target (Fig. 1b). Indeed, according to this hypothesis, lesions that interrupted such an instructive signal might in fact occlude the effects of altering auditory feedback (Fig. 1d). To test this hypothesis, we therefore compared the effects of deafening in intact adult zebra finches with the effects of deafening in birds that also received bilateral lesions of LMAN, the output nucleus of the anterior forebrain (Fig. 2b).

In line with previous reports¹, we found that songs of normal adult zebra finches were stable over long periods of time, while songs of zebra finches that had been deafened in adulthood gradually deteriorated over a period of weeks to months. Examples illustrating the range of song deterioration following deafening are shown in Fig. 3A,C. The effects of deafening included alteration of the structure of individual syllables, as well as changes to the overall temporal organization of the song. In marked contrast, songs of birds that received AFP lesions at the same time as they were deafened remained essentially unchanged (for example, Fig. 3B,D), and in most cases the subtle changes to syllable structure or syllable sequencing were comparable to those observed in control birds. Furthermore, lesions did not simply delay the effects of deafening; rather, they prevented changes in song for as long as the birds were followed (up to a year; see, for example, Fig. 3D, bottom panel).

The effects of experimental manipulations on song were characterized using two measures that focused separately on changes to individual syllables and on changes to the overall organization of song. These two aspects of song are thought to be controlled separately by RA (syllables) and HVC (syllable sequencing/temporal pattern)^{17,18}, and thus might be subject to independent changes. First, we used a subjective measure similar to that used in previous studies of birdsong to score (blind to the treatment of individual birds) how well syllables that were initially present in the bird's repertoire were preserved in later songs (see Methods). This measure was based on spectrographic representations of syllables that were presented individually, and therefore emphasized changes to the structure of syllables and discounted any changes to the sequence in which syllables were sung or the overall temporal pattern of song.

A summary of the changes in the structure of syllables for the different experimental groups is shown in Fig. 4a. For deafened birds, changes in syllables varied between individuals, and in some cases were no greater than those observed in control birds that retained normal hearing. Despite this variability, the songs of deafened birds deteriorated significantly relative to those of controls. Indeed, on average, the degree of similarity to the birds' original songs was not significantly greater than the similarity between songs randomly selected from unrelated birds. In contrast, the songs of birds that received LMAN lesions before deafening did not change more than those of control birds. Furthermore, in each instance where there were recordings from pairs of brothers, which controlled for some potential sources of variability, the syllables of the birds lesioned before deafening were more stable than those of their deafened but unlesioned brothers. Finally, 'sham' lesions, outside of the song system, did not prevent deafening induced deterioration of song, indicating that the effects of lesions were not due to non-specific damage to the brain. Thus, the absence of LMAN prevented the significant syllable deterioration that normally results from deafening in adult zebra finches.

Because of the deterioration of syllables in deafened birds, it generally was not possible to assess changes to the sequences in which syllables were sung. We therefore used a cross-correlation measure that was independent of the identification of syllables to characterize changes in the overall temporal pattern of song (see Methods). To illustrate this independence, we applied both

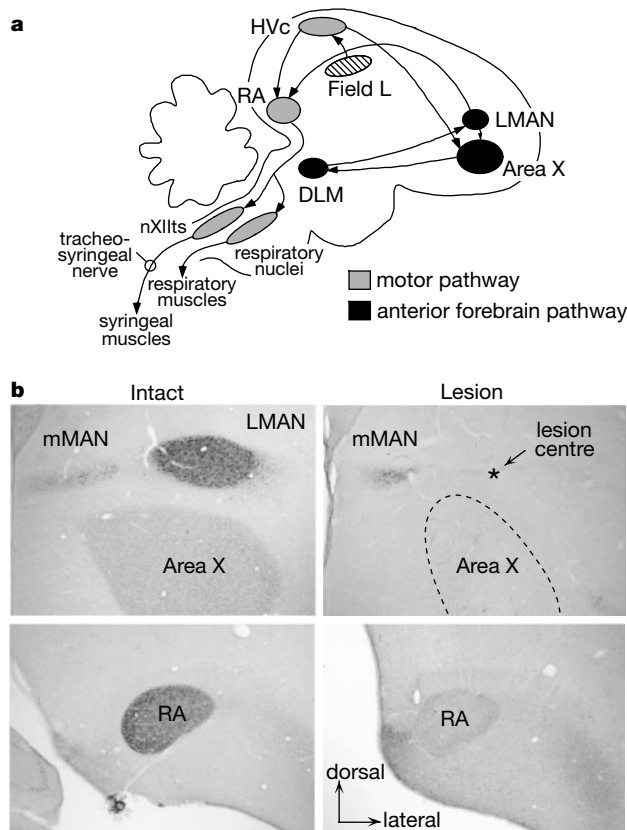


Figure 2 Song system nuclei in intact and lesioned birds. **a**, Schematic of connections within the song system. The 'motor pathway' (shaded) is required throughout life for song production⁹. The 'anterior forebrain pathway' (solid) is required for song learning, but not for adult song production. Field L and related areas provide auditory input to the song system. **b**, Transverse, CGRP-labelled²⁹ sections through the right anterior forebrain (top) and archistriatum (bottom). In an intact bird (left), LMAN was labelled along with its axonal projections into RA and Area X. The medial nucleus of the anterior neostriatum (mMAN) was also labelled. In a lesioned bird (right), nucleus LMAN was removed, with a corresponding loss of label in RA and Area X. Arrows, 250 μ m.

measures of song stability to two birds that were recorded immediately before and after cutting the tracheosyringeal nerve (Fig. 2a), which is essential for much of the structure of individual syllables, but not for the timing of patterned expiration during song¹⁹. The measures indicated a complete deterioration of syllable structure (Fig. 4a, 'nerve cut'), while the overall temporal pattern of song was well preserved (Fig. 4b, 'nerve cut'), demonstrating that changes to these two aspects of song were dissociable.

Despite this dissociability, changes to the temporal pattern of song for the experimental birds paralleled changes to syllable structure; deafening caused a significant deterioration of the temporal pattern of song relative to controls, and this deterioration was

prevented by lesions of LMAN but not by sham lesions (Fig. 4b). Thus, lesions of LMAN block deafening-induced vocal plasticity in adult zebra finches, not only of syllable structure, but also of the overall temporal pattern of song.

The results are consistent with the hypothesis outlined above (Fig. 1c), that disruption of auditory feedback in adult birds leads to the generation of an instructive signal that actively drives changes in song, and that this signal either arises within or is conveyed through the AFP. This possibility is supported both by the song-selective properties of AFP neurons, and by the recent finding that this circuit is active during singing, even in deaf birds¹⁵. It is critical to our experiment that, in deafened birds, song production is initially

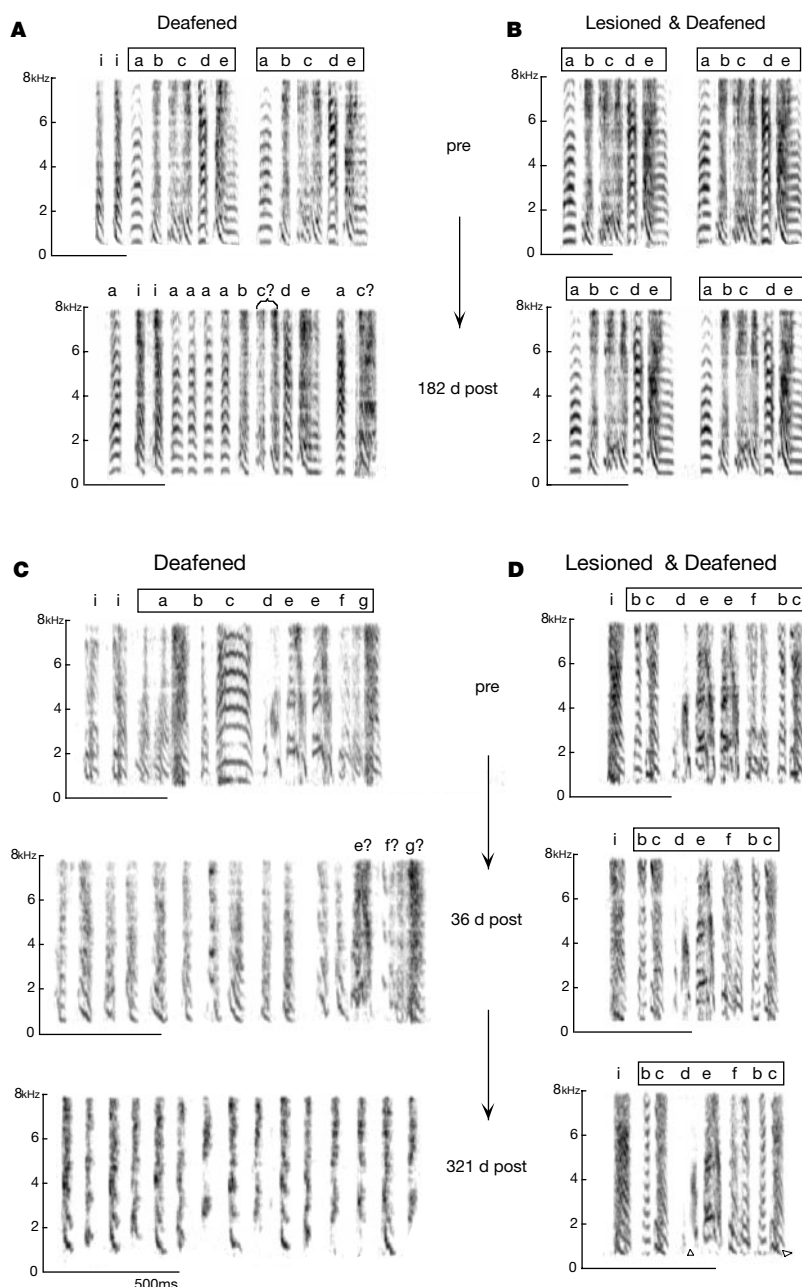


Figure 3 Changes to song following deafening in intact and lesioned birds. Shown are spectrograms illustrating changes to song for 2 deafened birds (**A** and **C**) and their 2 brothers, which additionally received LMAN lesions (**B** and **D**). Each bird initially sang a fixed repertoire of syllables (labelled with letters) in a stereotyped sequence or 'motif' (boxes). One of the smaller effects of deafening is shown in **A**: gradual loss or deterioration of some syllables (for example, c) and introduction of abnormal sequences (for example, repeated a's). **C**, One of the larger effects of deafening: relatively rapid loss of identifiable

syllables and eventual deterioration of song into a series of short, noisy notes. In contrast, songs of the lesioned brothers (**B** and **D**) recognizably retained all original syllables and sequencing. The largest change for these birds was loss of repetition of one syllable (e in panel **D**). Songs for all groups of birds also speeded up, as previously observed in controls³⁰ and following LMAN lesions²¹. This is apparent from the time scale bars (500 ms).

unchanged but song feedback to the nervous system is altered, that is, that the deafening is a 'pure' sensory manipulation. The data therefore also provide insight into two previously reported effects of adult lesions of LMAN: the prevention of incorporation of new syllables into songs of birds experimentally manipulated to undergo late learning²⁰, and the prevention of gradual changes to the abnormal songs of adult birds whose motor production has been disrupted by denervation of the vocal musculature²¹; this motor disruption also results in altered sensory feedback, and is thus a 'mixed', sensorimotor, manipulation. Our results with simple elimination of auditory feedback suggest that in both of the previous, more complex experimental situations, LMAN lesions may also act by eliminating signals from the AFP to the motor pathway about the mismatch between sensory feedback and the stored target song, thereby eliminating any impetus for change in vocal output (Fig. 1d).

An alternative (and not mutually exclusive) hypothesis about the role of the AFP is that neural or trophic inputs from this circuit are 'permissive' for plasticity in the motor pathway. According to this hypothesis, the instructive signal driving changes in song might arise elsewhere, but without the integrity of the AFP would be unable to bring about changes in song. This possibility is consistent with LMAN's known trophic support of RA early in development^{22,23}. Assessing whether the AFP is instructive, permissive or both will ultimately require characterization of the nature of signals generated in the AFP by alteration of auditory feedback.

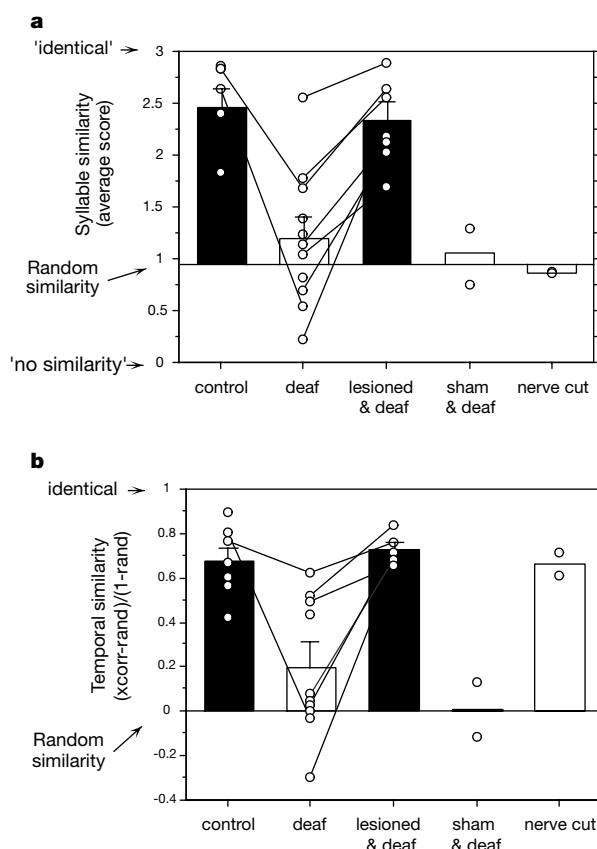


Figure 4 Summary of song stability following deafening of intact or LMAN lesioned birds. **a**, Syllable stability. Each point corresponds to an individual bird and shows the average similarity (see Methods) between syllables from baseline songs and those from songs recorded at least six months later, except for nerve-cut birds, where song changes were assessed within 1 week. Lines connect points corresponding to brothers. Bars show means and standard errors for each group. **b**, Temporal stability. Each point shows the temporal similarity (see Methods) between baseline songs and songs recorded after the indicated manipulations.

Regardless of the means by which AFP lesions prevent song plasticity, our study indicates that the changes in song that occur following interruption of auditory feedback are the result of an active process. This did not need to be the case. For example, the slow changes in song that occur following hearing loss could have reflected an uncorrected 'drift' in motor output that resulted from passive processes such as cell death or neurogenesis in the motor pathway²⁴, or from changes in the efficacy of vocal musculature. Since song changes are prevented by lesions of LMAN, however, this seems unlikely. Presumably, in adult birds, the AFP normally participates in the correction of any changes in song that result from such passive processes or from external damage to the motor pathway. Our results further demonstrate that in healthy adult birds such changes to the motor pathway are remarkably small; although removal of auditory feedback can lead to deterioration of song, auditory feedback is not required (in birds with AFP lesions) for song to remain stable.

Finally, the results are consistent with the AFP playing the same role in adults as has previously been hypothesized for juveniles: the auditory feedback-based evaluation and adaptive modification of the bird's own song. Differences between the effects of lesions in juvenile and adult zebra finches do not necessarily reflect a changing function of the AFP but rather a different state of motor competence; in juveniles, lesions disrupt an instructive signal so that normal song learning cannot progress, while in adults that have completed song learning the presumptive signal for change is small so that lesions have little effect. In the adult case, the effects of lesions are only revealed under conditions (such as disruption of auditory feedback) where a large instructive signal to alter song is elicited.

This 'covert' contribution of the AFP to adult plasticity may have parallels to mammalian systems, where damage to similar reciprocal cortical-basal ganglia circuits can impair procedural learning while having little effect on previously learned performance^{25–27}. In particular, since speech, like song, is an apparently stable behaviour that can nevertheless change in response to altered auditory feedback^{2,6}, our results suggest that it may be revealing to examine the functioning of cortical-basal ganglia circuits during the modification of speech; there is already evidence that such circuits are differentially active during production of speech in a second language versus a native language²⁸. Since the AFP is a discrete basal ganglia–forebrain circuit specialized for one well defined behaviour, it may prove a particularly tractable system for elucidating the signals present in these structures and the role that they play in the learning and modification of sequenced motor acts. □

Methods

Animals

Zebra finches (*Taeniopygia guttata*) were raised in individual breeding cages with their parents and siblings until approximately 60 days of age when they were transferred to all-male group cages. Preliminary experiments suggested that there were greater effects of deafening on young adult birds (90–140 days of age) than on older birds (>200 days of age). For this study, baseline recordings and experimental manipulations were therefore restricted to birds that were 98–140 days of age, except for one pair of brothers (one lesioned and one intact) that were deafened at 204 days of age. These two birds were included in Fig. 4, but not in statistical analyses. Where possible, we studied pairs of brothers that were housed together, experienced manipulations at the same ages, and had similar songs by virtue of exposure to each other and the same adult tutor. There was a significant correlation between syllable stability (see below) for deafened birds and syllable stability for their lesioned brothers, indicating that the identity of brothers indeed accounted for some of the variability in song stability (see also Fig. 4, where data that derive from brothers are connected by lines).

Lesions

Birds were deafened by bilateral cochlear removal⁷. For some birds, LMAN was electrolytically lesioned 1–5 days before deafening. Lesions were stereotactically targeted at LMAN, and evaluated in tissue labelled with an antibody to CGRP (Fig. 2)²⁹. For all but one of the nominally lesioned birds, the percentage of LMAN that was removed bilaterally ranged from 70% to 100%. The remaining bird had only a 37% lesion, but it included the region where RA projecting axons leave LMAN, and almost all CGRP labelling in RA was eliminated. 'Sham' lesions were entirely anterior and dorsal to LMAN.

Analysis of changes to song

To characterize changes to the structure of individual syllables, we used a subjective scoring procedure similar to that employed in previous studies of birdsong^{16–20}. Observers who were blind to the experimental manipulation of each bird scored the similarity between spectrographic representations of syllables from songs recorded before any manipulation and those from later songs. For each baseline syllable that was initially present in the repertoire of a bird (5–12 distinct syllables per bird), observers identified the most similar syllable present in songs recorded at later dates. Observers additionally judged the degree of similarity between each baseline syllable and its best match on a scale of 0 (no similarity) to 3 (identical). As a measure of the maximum expected deterioration of syllables, we determined the random similarity between syllables of songs from unrelated birds ($n = 12$ pairs of songs). This random similarity averaged 0.9, indicating some generic similarity between the syllables of unrelated zebra finches. Scores were averaged across 3 observers (mean r^2 for pairwise correlation of observers' scores, 0.81). All songs were recorded from birds isolated in sound attenuating boxes and were therefore 'undirected'. Significance of differences between groups was based on a criterion of $P < 0.05$ using analysis of variance (ANOVA) and post-hoc Fischer's Protected Least Significant Difference tests.

To assess temporal stability, songs were represented by the timing of syllable production, while the spectral structure of individual syllables was ignored. For each bird, the timing pattern representing the most common motif before any experimental manipulation was identified. We then searched songs from later recording sessions for the pattern of syllables that provided the closest temporal match to the initial motif. At each offset (in increments of 5 ms) between motif and song, we quantified the temporal overlap as the average of the percentage overlap of syllables and the percentage overlap of intervals. For each bird, the maximal overlap was calculated for 10 songs and then averaged to provide a measure ('xcorr') of how well the temporal pattern of song was conserved following experimental manipulation. Because the delivery of some songs speeded up 'proportionately' over time (that is, with no apparent change in the relative durations of notes and intervals), we allowed for proportional changes in the temporal pattern of the song (ranging from 75% to 110% of initial song duration) in searching for the maximal overlap. To control partially for varying complexity of different birds' motifs, we calculated the maximal overlaps between each bird's motif and randomly selected songs from unrelated birds ('rand'). The temporal stability of song was then expressed as a normalized value ranging from 0 (indicating no more preservation of temporal pattern than random) to 1 (indicating perfect preservation of temporal pattern). The baseline motif for one pair of brothers had a very simple temporal pattern (3 syllables) and consequently showed a high degree of temporal similarity to all songs (including those from unrelated birds). This pair of birds was therefore excluded from further analysis of changes in temporal pattern.

Received 12 December 1999; accepted 1 February 2000.

- Nordeen, K. W. & Nordeen, E. J. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57**, 58–66 (1992).
- Cowie, R. & Douglas-Cowie, E. *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* 1–304 (Mouton de Gruyter, Berlin, 1992).
- Marler, P. A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* **71**, 1–25 (1970).
- Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
- Leonardo, A. & Konishi, M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399**, 466–470 (1999).
- Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
- Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
- Bottjer, S. W. & Johnson, F. Circuits, hormones, and learning: vocal behavior in songbirds. *J. Neurobiol.* **33**, 602–618 (1997).
- Nottebohm, F., Stokes, T. M. & Leonard, C. M. Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* **165**, 457–486 (1976).
- Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901–903 (1984).
- Sohrabji, F., Nordeen, E. J. & Nordeen, K. W. Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behav. Neural Biol.* **53**, 51–63 (1990).
- Scharff, C. & Nottebohm, F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11**, 2896–2913 (1991).
- Doupe, A. J. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.* **17**, 1147–1167 (1997).
- Solis, M. M. & Doupe, A. J. Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. *J. Neurosci.* **19**, 4559–4584 (1999).
- Hessler, N. A. & Doupe, A. J. Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches. *J. Neurosci.* **19**, 10461–10481 (1999).
- Nordeen, K. W. & Nordeen, E. J. Long-term maintenance of song in adult zebra finches is not affected by lesions of a forebrain region involved in song learning. *Behav. Neural Biol.* **59**, 79–82 (1993).
- Vu, E. T., Mazurek, M. E. & Kuo, Y. C. Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.* **14**, 6924–6934 (1994).
- Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
- Vicario, D. S. Contributions of syringeal muscles to respiration and vocalization in the zebra finch. *J. Neurobiol.* **22**, 63–73 (1991).
- Morrison, R. G. & Nottebohm, F. Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *J. Neurobiol.* **24**, 1045–1064 (1993).
- Williams, H. & Mehta, N. Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *J. Neurobiol.* **39**, 14–28 (1999).

- Johnson, F. & Bottjer, S. W. Afferent influences on cell death and birth during development of a cortical nucleus necessary for learned vocal behavior in zebra finches. *Development* **120**, 13–24 (1994).
- Akutagawa, E. & Konishi, M. Two separate areas of the brain differentially guide the development of a song control nucleus in the zebra finch. *Proc. Natl Acad. Sci. USA* **91**, 12413–12417 (1994).
- Kirn, J. R. & Nottebohm, F. Direct evidence for loss and replacement of projection neurons in adult canary brain. *J. Neurosci.* **13**, 1654–1663 (1993).
- Gabyiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
- Knowlton, B. J., Mangels, J. A. & Squire, L. R. A neostriatal habit learning system in humans. *Science* **273**, 1399–1402 (1996).
- Nakamura, K., Sakai, K. & Hikosaka, O. Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. *J. Neurophys.* **82**, 1063–1068 (1999).
- Klein, D., Zatorre, R. J., Milner, B., Meyer, E. & Evans, A. C. Left putaminal activation when speaking a second language: evidence from PET. *Neuroreport* **5**, 2295–2297 (1994).
- Bottjer, S. W., Roselinsky, H. & Tran, N. B. Sex differences in neuropeptide staining of song-control nuclei in zebra finch brains. *Brain Behav. Evol.* **50**, 284–303 (1997).
- Arnold, A. P. The effects of castration on song development in zebra finches (*Poephila guttata*). *J. Exp. Zool.* **191**, 261–278 (1975).

Acknowledgements

We thank M. Stryker, M. Churchland, T. Troyer and K. T. Moortgat for comments on the manuscript, and A. Arteseros, G. Carrillo and A. Tam for technical assistance. This work was supported by a Burroughs Wellcome Fund fellowship of the Life Sciences Research Foundation (M.S.B.), and by the John Merck Fund, the EILB Foundation and the National Institutes of Health (A.J.D.).

Correspondence and requests for materials should be addressed to M.S.B. (e-mail: mshb@phy.ucsf.edu).

SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*

Sarah J. Liljegren*, Gary S. Ditta*, Yuval Eshed†, Beth Savidge*‡, John L. Bowman† & Martin F. Yanofsky*

* Section of Cell and Developmental Biology, University of California at San Diego, La Jolla, California 92093-0116, USA

† Section of Plant Biology, University of California at Davis, Davis, California 95616, USA

The fruit, which mediates the maturation and dispersal of seeds, is a complex structure unique to flowering plants. Seed dispersal in plants such as *Arabidopsis* occurs by a process called fruit dehiscence, or pod shatter. Few studies^{1–3} have focused on identifying genes that regulate this process, in spite of the agronomic value of controlling seed dispersal in crop plants such as canola^{4,5}. Here we show that the closely related *SHATTERPROOF* (*SHP1*) and *SHATTERPROOF2* (*SHP2*) MADS-box genes are required for fruit dehiscence in *Arabidopsis*. Moreover, *SHP1* and *SHP2* are functionally redundant, as neither single mutant displays a novel phenotype. Our studies of *shp1 shp2* fruit, and of plants constitutively expressing *SHP1* and *SHP2*, show that these two genes control dehiscence zone differentiation and promote the lignification of adjacent cells. Our results indicate that further analysis of the molecular events underlying fruit dehiscence may allow genetic manipulation of pod shatter in crop plants.

The MADS-box gene family encodes transcriptional regulators involved in diverse aspects of plant development, and phylogenetic and functional studies show extensive redundancy between family members^{6–9}. *SHP1* and *SHP2* (previously known as *AGL1* and *AGL5*) were two strong candidates for functional redundancy, as they share 87% identity at the amino-acid sequence level and show almost identical expression patterns in developing *Arabidopsis* fruit (refs 10–12; and C. Ferrándiz, Y.E., J.L.B. and M.F.Y., unpublished results).

‡ Present address: Calgene, Davis, California 95616, USA.

Arabidopsis is typical of the more than 3,000 species of the Brassicaceae, producing dry dehiscent fruit in which two carpel valves are joined to the replum, a thin structure that separates the valves (Fig. 1)^{13–17}. At the valve–replum boundary (valve margin), a

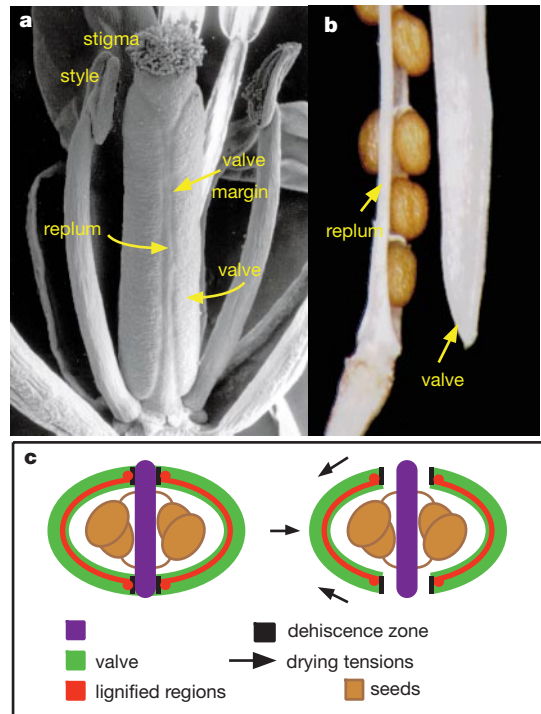


Figure 1 *Arabidopsis* fruit structure and dehiscence. **a, b**, Wild-type gynoecium at fertilization (**a**) and during fruit dehiscence (**b**). **c**, Proposed factors contributing to fruit dehiscence. Before dehiscence, small patches of valve margin cells immediately adjacent to the dehiscence zone become lignified, as well as an internal valve cell layer. As the fruit dries, the thin-walled cells of the valves shrink, creating tension within the rigid lignified regions. Once middle lamellar breakdown between dehiscence zone cells occurs, this tension contributes to valve separation from the replum.

narrow band of cells develops into the dehiscence zone¹³. Late in fruit development (stage 18/19), breakdown of the middle lamella between dehiscence zone cells leads to detachment of the valve from the replum, allowing seed dispersal (Fig. 1b)^{2,13,18}. Lignification of valve margin cells adjacent to the dehiscence zone and of an internal valve cell layer is proposed to contribute to pod shatter (Fig. 1c)¹⁴. Within the developing gynoecium, *SHP1* and *SHP2* are expressed in four medial domains which become narrow stripes (stage 9) before the valve margin is distinct (stage 12). This expression pattern at the valve margin is maintained after fertilization, indicating that *SHP1* and *SHP2* may function both to specify the valve margin and to direct dehiscence zone development in the mature fruit.

Using a polymerase chain reaction (PCR)-based screen of a T-DNA insertional collection, we identified a line containing a disruption in the *SHP1* gene. Subsequent RNA blot analysis confirmed our isolation of a putative *shp1* loss-of-function allele (data not shown). The *SHP2* gene was previously targeted for knockout by homologous recombination¹⁹. Although both single mutants have putative loss-of-function alleles, *shp1-1* and *shp2-1* fruit show no detectable differences from wild-type fruit (data not shown). In contrast, *shp1 shp2* double mutants have a striking phenotype, as the mature fruit fail to dehisce. Through *shp1* and *shp2* mutagenesis screens for plants producing indehiscent fruit, we have identified additional mutant alleles of *SHP2* and *SHP1*, respectively.

Phenotypic differences between the valve margins of *shp1 shp2* and wild-type fruit are first detected after fertilization. At this point, valve margin cells (about four cell files) in wild-type fruit expand more slowly than valve cells, resulting in a noticeable constriction of the valve margin¹³. Scanning electron microscopy and transverse sections of *shp1 shp2* fruit (stage 14) show less constriction of the valve margins than wild-type fruit (data not shown). Later in wild-type fruit development (stage 16/17), valve margin cells (two cell files) closest to the replum form the dehiscence zone, and separation between dehiscence zone cells begins to occur in mature fruit (stage 18)¹³. Scanning electron microscopy reveals the absence of dehiscence zones in mature *shp1 shp2* fruit at stage 18 (Fig. 2).

As lignification within the fruit may contribute to pod shatter¹⁴, we examined the lignification pattern of *shp1 shp2* fruit using phloroglucinol, a lignin-specific histological stain (Fig. 3). Whereas

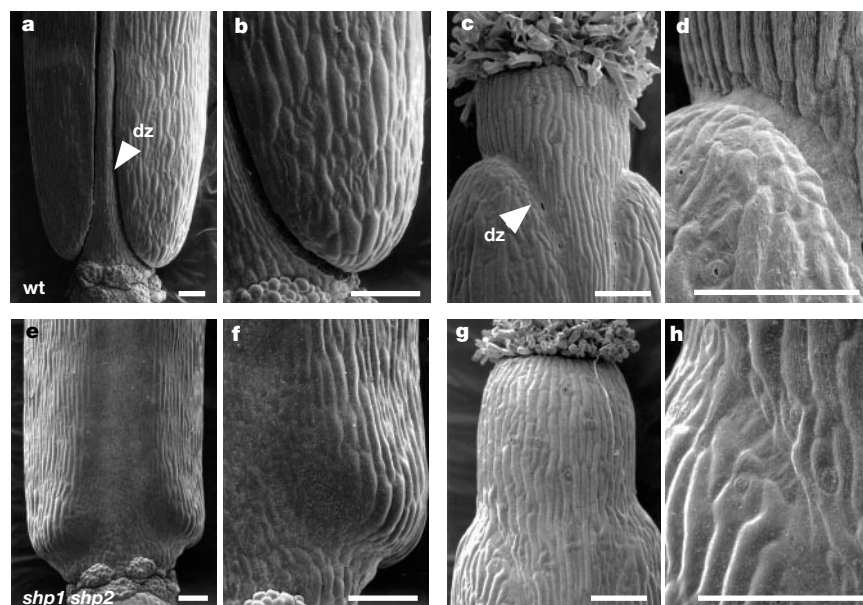


Figure 2 *SHP1* and *SHP2* act redundantly and are required for dehiscence zone differentiation. **a–d**, Wild-type (wt) fruit (stage 18) show well developed dehiscence zones (dz) and separation of the valves from the replum is already apparent at the fruit base

(**a, b**). **e–h**, Dehiscence zones are absent in *shp1 shp2* fruit (stage 18). The valve margins of *shp1 shp2* fruit show far less definition than those of wild-type fruit, especially at the fruit base (**e, f**). Scale bars, 100 μ m.

wild-type fruit (stage 17) show lignification of valve margin cells adjacent to the dehiscence zone throughout the fruit (Fig. 3a, c), we observed a reduction in valve margin cell lignification in *shp1 shp2* fruit. These alterations are regional, as there was no valve margin lignification at the base of *shp1 shp2* fruit (Fig. 3b), but there were lignified valve margin cells toward the fruit apices (Fig. 3d). Lignification of an internal valve cell layer, which occurs slightly later in fruit development¹³, is not affected in *shp1 shp2* mutants (Fig. 3b, d).

As our results indicate that *SHP1* and *SHP2* regulate valve margin development, we used two molecular markers derived from gene and enhancer trap screens^{20,21} to monitor the cellular differentiation of wild-type and *shp1 shp2* valve margins and to identify possible downstream targets of *SHP1* and *SHP2*. Plants containing the valve margin molecular markers GT140 (ref. 20) and YJ36, which have distinct expression profiles, were crossed to *shp1 shp2* plants. Both markers show very different expression patterns in *shp1 shp2* fruit, demonstrating the altered fate of *shp1 shp2* valve margins (Fig. 4). In wild-type fruit, GT140 is expressed in stripes at the valve margin (Fig. 4a) and in a diffuse domain at the valve base (data not shown). Transverse sections of *shp1 shp2* fruit (stage 17) show that this marker is largely absent from cells at the valve margin (Fig. 4b), although a low level of expression remains at the base of the valves (data not shown). YJ36 is expressed in the outer and inner epidermal cells of the valve margin, the septum and the nectaries of wild-type fruit (Fig. 4c, and data not shown). The valve margin expression domains of this marker are largely absent in *shp1 shp2* fruit whereas the other expression domains remain unchanged (Fig. 4d).

These loss-of-function analyses show that *SHP1* and *SHP2* encode redundant proteins required for proper development of the *Arabidopsis* fruit valve margin, including differentiation of dehiscence zone cells and lignification of valve margin cells adjacent to the dehiscence zone. The altered expression patterns of both GT140 and YJ36 in *shp1 shp2* fruit suggest that their genes act downstream of, and are positively regulated by, *SHP1* and *SHP2*. As these genes are excellent candidates for transcriptional targets of *SHP1* and *SHP2*, we are currently investigating their identity and looking for other downstream targets of *SHP1* and *SHP2* at the valve margin.

No other phenotypes were observed in the *shp1 shp2* double mutant, even though *SHP1* and *SHP2* are also expressed in developing ovules, the septum, the nectaries and the style^{11,12}. *SHP1* and *SHP2* activities in these regions may be functionally redundant with other genes, such as the MADS-box genes *AGAMOUS* (*AG*) and *AGL11*, which are closely related to *SHP1* and *SHP2* and are also expressed in developing ovules^{22–24}. Further analysis of *SHP1* and *SHP2* function in ovule development may have to await identification of an *agl11* loss-of-function allele and characterization of the *agl11 shp1 shp2* triple mutant.

To complement our loss-of-function analyses of *SHP1* and *SHP2*, we also examined transgenic gain-of-function plants that constitutively express either *SHP1* or *SHP2* under control of the cauliflower mosaic virus 35S promoter²⁵ (35S::*SHP1* and 35S::*SHP2*). Flowers from plants constitutively expressing *SHP1* or *SHP2* show a phenotype similar to flowers from plants constitutively expressing the MADS-box gene *AG*²⁶. Carpeloid sepals and staminoid petals are present in the first and second whorls, respectively, of 35S::*SHP1* and 35S::*SHP2* flowers (data not shown). In addition, 35S::*SHP1* and 35S::*SHP2* fruit show a range of phenotypes, which are most consistent in plants carrying both the *SHP1* and *SHP2* transgenes (Fig. 5). 35S::*SHP1* 35S::*SHP2* fruit are nearly wild-type in length, but have a significantly smaller diameter. The valves of 35S::*SHP1* 35S::*SHP2* fruit (stage 17) have a yellowish-green appearance compared to the dark green colour of wild-type valves, and, before seed maturity, large tears can occur in the valve regions because of seed crowding (Fig. 5a).

Scanning electron microscopy shows that the valves of 35S::*SHP1* 35S::*SHP2* fruit (stage 17) differ from wild-type fruit valves (Fig. 5b). Guard cells and their associated subsidiary cells make up a large fraction of the outer epidermis of wild-type fruit valves (Fig. 2b), but these cells are not apparent in 35S::*SHP1* 35S::*SHP2* valves. Instead, 35S::*SHP1* 35S::*SHP2* valve outer epidermal cells appear homogeneously long and slender. Furthermore, the valve inner epidermis of 35S::*SHP1* 35S::*SHP2* fruit contains about twice as many cells as seen in wild-type fruit (Fig. 5e, f).

Our loss-of-function studies indicate that *SHP1* and *SHP2* promote lignification of a subset of valve margin cells, so we analysed the lignification patterns of fruit that ectopically express

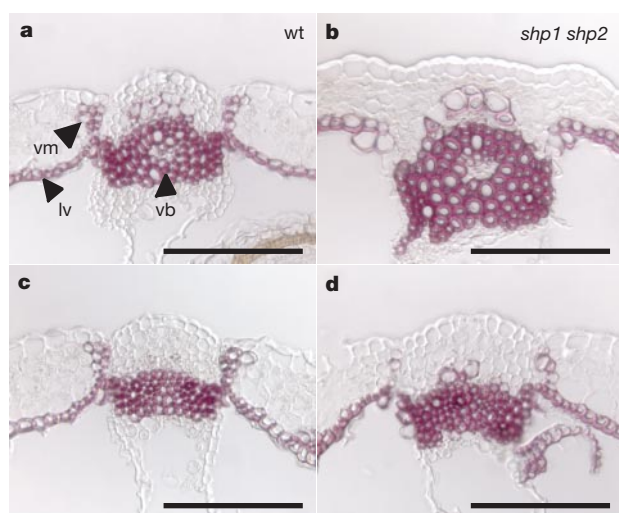


Figure 3 *SHP1* and *SHP2* promote valve margin lignification. Transverse sections of wild-type (a, c) and *shp1 shp2* (b, d) fruit (stage 17) were stained with phloroglucinol. Small patches of valve margin (vm) cells immediately adjacent to the dehiscence zones are lignified throughout wild-type fruit (a, c), whereas these lignified cell patches are not seen at the bases of *shp1 shp2* fruit (b) and are reduced elsewhere in *shp1 shp2* fruit (d). Lignification of the vascular bundles (vb) or of an internal valve cell layer (lv) is not affected in *shp1 shp2* fruit. Scale bars, 100 μm.

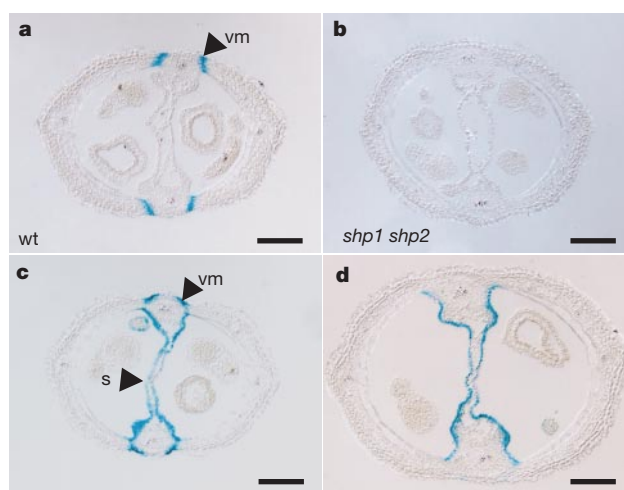


Figure 4 *SHP1* and *SHP2* regulate the expression of valve margin molecular markers. a, b, In wild-type fruit (a), the GT140 marker is expressed in stripes at the valve margin (vm); it is largely absent in *shp1 shp2* fruit (b). c, d, The YJ36 marker is expressed on the outer and inner surfaces of the wild-type valve margin (c) but not in *shp1 shp2* (d) fruit (stage 17). YJ36 expression is also found within the septum (s) of both wild-type and *shp1 shp2* fruit. Scale bars, 100 μm.

SHP1, *SHP2* or both genes. Although the lignification pattern of 35S::*SHP2* fruit appears identical to that of wild-type fruit, 35S::*SHP1* and 35S::*SHP1* 35S::*SHP2* fruit (stage 17) show ectopic lignification (Fig. 5d, and data not shown) of the valve mesophyll layers, with the most extensive lignification apparent in 35S::*SHP1* 35S::*SHP2* fruit. Cells at positions corresponding to the mesophyll layers are lignified at the valve margins of wild-type fruit (Fig. 5c), so ectopic lignification of the 35S::*SHP1* 35S::*SHP2* valve mesophyll layers suggests that these layers have acquired a lignified valve margin cell identity.

As the GT140 and YJ36 molecular markers demonstrated the altered fate of *shp1 shp2* valve margins, plants containing these markers were crossed to 35S::*SHP1* 35S::*SHP2* plants. GT140 is expressed in stripes at the valve margin of wild-type fruit (Fig. 4a), but in mature 35S::*SHP1* 35S::*SHP2* fruit (stage 16/17) expression of this marker expands throughout the valves (Fig. 5g). Expansion of this marker is consistent with transformation of these valves to a valve margin fate, and with positive regulation of the corresponding gene by *SHP1* and *SHP2*. The expression profiles of YJ36 in wild-type and 35S::*SHP1* 35S::*SHP2* fruit are the same (data not shown), indicating that the corresponding gene is not involved in the transformed valve margin cell identity of 35S::*SHP1* 35S::*SHP2* fruit valves.

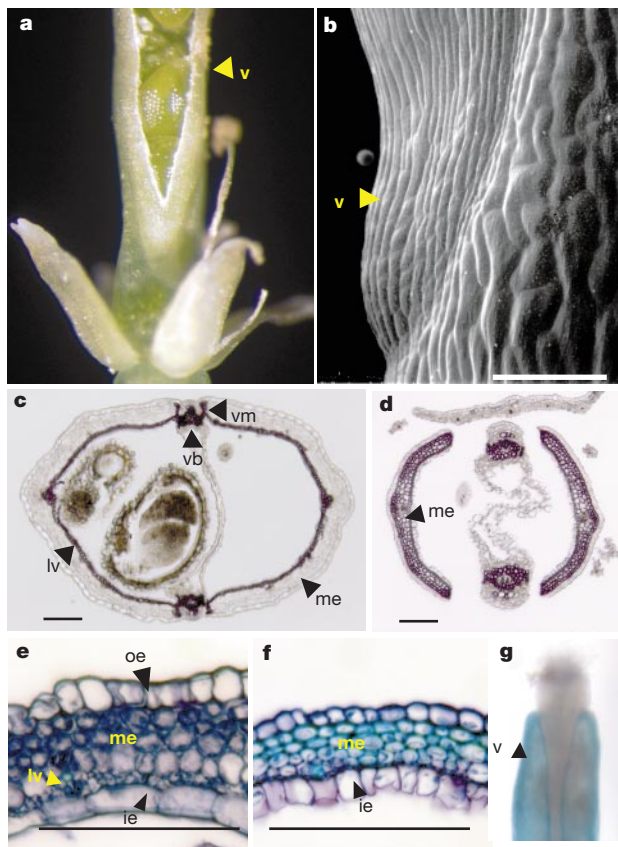


Figure 5 Ectopic expression of *SHP1* and *SHP2* interferes with valve development. **a**, Seed crowding can cause the valves (v) of 35S::*SHP1* 35S::*SHP2* fruit (stage 17) to tear open before fruit dehiscence. **b**, Guard cell differentiation does not occur in the outer epidermis (oe) of 35S::*SHP1* 35S::*SHP2* fruit valves (see wild-type fruit valve, Fig. 2b). Instead, the outer epidermis consists of a homogeneous population of long, slender cells. **c–f**, Transverse sections of wild-type (**c**, **e**; stage 17 and 16, respectively) and 35S::*SHP1* 35S::*SHP2* (**d**, **f**) fruit (stage 17) stained with phloroglucinol (**c**, **d**) or toluidine blue (**e**, **f**). The valve mesophyll (me) cell layers are ectopically lignified (**d**) and the valve inner epidermis (ie) has about twice as many cells (**f**) in 35S::*SHP1* 35S::*SHP2* fruit. **g**, Expression of the GT140 valve margin marker expands throughout the valves of 35S::*SHP1* 35S::*SHP2* fruit (stage 16). Scale bars, 100 μ m.

These observations indicate that constitutive expression of *SHP1* and *SHP2* throughout the fruit alters development of the valve outer epidermis, the mesophyll layers and the inner epidermis, in a manner consistent with transformation of the valves to a valve margin fate, specifically to that of the valve margin cells adjacent to the dehiscence zone. As valve margin cells do not expand at the same rate as cells throughout the remainder of the valve¹³, these transformations may explain the smaller diameter of 35S::*SHP1* 35S::*SHP2* fruit. These gain-of-function analyses suggest that neither *SHP1* nor *SHP2* alone is sufficient to specify a valve margin fate throughout the fruit valve, as the severity of the observed phenotypes is considerably stronger when both transgenes are present. However, it is probable that these two genes do act redundantly and that the severity of the phenotypes is simply a reflection of the higher levels of *SHP* expression in the doubly transgenic plants.

Remarkable phenotypic similarities are apparent between 35S::*SHP1* 35S::*SHP2* fruit and fruit carrying mutant alleles of the *FRUITFULL* (*FUL*) MADS-box gene, although *ful* mutant siliques are much shorter than those of wild-type fruit, and the *ful* mutants exhibit additional phenotypes not seen in 35S::*SHP1* 35S::*SHP2* fruit²⁷. In the developing fruit, *FUL* is expressed throughout the valves in a complementary pattern to that of *SHP1/SHP2* at the valve margin, and is required for valve cell differentiation and expansion after fertilization^{27,28}. Like 35S::*SHP1* 35S::*SHP2* fruit, *ful* fruit exhibit substantial seed crowding because of a lack of lateral valve expansion, differentiation defects of the valve outer epidermis, ectopic lignification of the valve mesophyll layers and a significantly increased number of valve inner epidermal cells (ref. 27; and S.J.L., C. Ferrández and M.F.Y., unpublished results). Similarities also exist between the *shp1 shp2* loss-of-function and 35S::*FUL* gain-of-function phenotypes. Like the *shp1 shp2* double mutants, plants that constitutively express *FUL* also produce indehiscent fruit (C. Ferrández, S.J.L. and M.F.Y., unpublished results). These parallels suggest that the *SHP* and *FUL* MADS-box genes may interact antagonistically during valve margin development, and it will be intriguing to explore these possible interactions through expression analyses, construction of the *shp1 shp2 ful* triple mutant and the use of molecular markers.

Arabidopsis is closely related to important oilseed crop plants such as canola (*Brassica napus*, *B. rapa*), where pod shatter causes average annual yield losses of 20% and up to 50% under adverse weather conditions^{4,5}. Our studies of *Arabidopsis* fruit dehiscence suggest several molecular strategies to reduce pod shatter in crop plants by genetic engineering. In addition, the availability of *Arabidopsis* mutants that fail to dehisce will allow further investigation of the cascade of gene activity that leads to formation of the valve margin and the programmed series of events within these cells that results in fruit dehiscence. □

Methods

Mutant screens

The *shp1-1* allele was obtained through a PCR-based screen of 5,600 T-DNA insertional lines, using a *SHP1*-specific oligonucleotide (oligo) (5'-CGTTATGAGGGGAAGAAACAA GTTATCATG-3') and an oligo from the T-DNA left border (5'-GATGCACTCGAAATC AGCCAATTTTAGAC-3'). The T-DNA insertion occurs in the large first intron of the *SHP1* locus. Additional alleles of *SHP1* and *SHP2* were obtained through ethyl methanesulphonate mutagenesis of *shp2-1* and *shp1-1* seed stocks, respectively. More than 4,000 *shp1-1* and 6,000 *shp2-1* M2 families were screened for plants with indehiscent fruit. From these screens, three independent alleles of *SHP2* and four independent alleles of *SHP1* were discovered. The *shp2-2* allele contains a nucleotide substitution at the splice acceptor site of the fourth intron. The same nucleotide substitution at codon 115 in both the *shp1-2* and *shp1-5* alleles changes a glutamine to a stop codon in the K domain, and the *shp1-4* allele contains a nucleotide substitution at the splice donor site of the fourth intron.

Mutant genotyping

Plants homozygous for the *shp1-1* and/or *shp2-1* alleles were identified by PCR genotyping as follows. Amplification with the oligo (5'-GTGACGGAAGGAGGTTGACG-3') and a T-DNA specific oligo (above) yields a 597-base pair (bp) product in plants containing the

shp1-1 allele; amplification using the oligos 5'-GTGACGGAAGGAGGGTTGACG-3' and 5'-GTCCTACTGATGAGTTGCTCACTAGG-3' yields an 871-bp product in plants with a wild-type allele of *SHP1*, and no product in plants homozygous for the *shp1-1* allele. Amplification using the oligos 5'-GAGGATAGAGAACAACACTACGAATCGTC-3' and 5'-CAGGTCAGTCAATAGATTCCCTAC-3' yields a 1.5-kb product in plants with a wild-type allele of *SHP2*, and a single 2.8-kb product in plants homozygous for the *shp2-1* allele.

Generation of transgenic plants

A full-length *SHP1* complementary DNA was created by fusing the *EcoRI* fragments of pCIT2241 and pCIT4219 (ref. 10). The *SHP1* cDNA was then cloned into the *Bam*HI site of pCNG18 (ref. 29), to place *SHP1* transcription under the control of the viral 35S promoter²⁵. A full-length *SHP2* cDNA was PCR amplified with the oligos 5'-GGAGATCTGAATTCAT CTTCCTCATCC-3' and 5'-CCGGTACCTCAACAAGTTG-CAGAGGTGGTGG TCTTGGTTGGAGGAATTCGATTCCGGTCAAG-3', using pCIT2242 (ref. 10) as template. After cloning this product into the TA vector (Invitrogen), a *Bgl*II/*Kpn*I fragment containing the *SHP2* cDNA was cloned into the plant transformation vector pMON530 (Monsanto). The resulting construct also places *SHP2* transcription under control of the 35S promoter. Transgenic plants were selected on kanamycin after *Agrobacterium*-mediated transformation. 35S::*SHP1* plants are in the Landsberg *erecta* ecotype, whereas 35S::*SHP2* plants are in the Columbia ecotype. 35S::*SHP1* and 35S::*SHP2* plants were crossed to each other; in the F1 generation, 35S::*SHP1* 35S::*SHP2* plants were identified by PCR genotyping using *SHP1* transgene-specific (5'-GAAGGTGGGAGTAGTCACGAC-3' and 5'-CGGAAGGAGGGTTGACGGCA-3') and *SHP2* transgene-specific (5'-GGTGGTCCGAGTAATGAAGTA-3' and 5'-TGGTCGGAGGGTTAACGGCG-3') oligos.

Scanning electron microscopy

Fruit from wild-type (Columbia ecotype), *shp1 shp2* mutants and 35S::*SHP1* 35S::*SHP2* plants were fixed for approximately 4 h at 25°C in FAA (50% ethanol, 5% glacial acetic acid, 3.7% formaldehyde) and prepared for scanning electron microscopy²⁷. Samples were examined in a Cambridge S360 scanning electron microscope using an accelerating voltage of 10 kV.

Histological staining

Tissue from wild-type (Columbia ecotype), *shp1 shp2* and 35S::*SHP1* 35S::*SHP2* plants was fixed, sectioned and stained with toluidine blue²⁶ with minor modifications. For lignin analyses, sections were stained for 2 min in a 2% phloroglucinol solution in 95% ethanol, then photographed in 50% hydrochloric acid.

Molecular marker analyses

The YJ36 enhancer trap line was generated by *Agrobacterium*-mediated transformation with the plasmid pOCA-28-15-991 (ref. 21). Transgenic plants containing YJ36 or GT140 (ref. 20) were crossed to *shp1 shp2* plants. Among the F2 population, plants with a *shp1 shp2* phenotype that also carried the respective molecular marker were selected for further study. F1 plants from marker crosses to 35S::*SHP1* 35S::*SHP2* plants were genotyped for both transgenes, and analysed further. For β -glucuronidase expression analyses, fruit from GT140, *shp1 shp2* GT140, 35S::*SHP1* 35S::*SHP2* GT140, YJ36, *shp1 shp2* YJ36 and 35S::*SHP1* 35S::*SHP2* YJ36 plants were fixed, sectioned and stained as described³⁰ with minor modifications.

Received 7 December 1999; accepted 16 February 2000.

- Jenkins, E. S. *et al.* Characterization of an mRNA encoding a polygalacturonase expressed during pod development in oilseed rape (*Brassica napus* L.). *J. Exp. Bot.* **47**, 111–115 (1996).
- Petersen, M. *et al.* Isolation and characterisation of a pod dehiscence zone-specific polygalacturonase from *Brassica napus*. *Plant Mol. Biol.* **31**, 517–527 (1996).
- Coupe, S. A., Taylor, J. E., Isaac, P. G. & Roberts, J. A. Identification and characterization of a proline-rich mRNA that accumulates during pod development in oilseed rape (*Brassica napus* L.). *Plant Mol. Biol.* **23**, 1223–1232 (1993).
- Child, R. D., Chauvaux, N., John, K., Ulvskov, P. & Onckelen, H. A. Ethylene biosynthesis in oilseed rape pods in relation to pod shatter. *J. Exp. Bot.* **49**, 829–838 (1998).
- MacLeod, J. In *Oilseed Rape Book* 107–119 (Cambridge Agricultural, Cambridge, 1981).
- Riechmann, J. L. & Meyerowitz, E. M. MADS domain proteins in plant development. *J. Biol. Chem.* **378**, 1079–1101 (1997).
- Purugganan, M. D. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *J. Mol. Evol.* **45**, 392–396 (1997).
- Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. & Smyth, D. R. Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119**, 721–743 (1993).
- Kempin, S. A., Savidge, B. & Yanofsky, M. F. Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267**, 522–525 (1995).
- Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5**, 484–495 (1991).
- Flanagan, C. A., Hu, Y. & Ma, H. Specific expression of *AGL1* MADS-box gene suggests regulatory functions in *Arabidopsis* gynoecium and ovule development. *Plant J.* **10**, 343–353 (1996).
- Savidge, B., Rounsley, S. D. & Yanofsky, M. F. Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell* **7**, 721–733 (1995).
- Spence, J. *Development of the Silique* of *Arabidopsis thaliana*. Thesis, Univ. Durham (1992).
- Spence, J., Vercher, Y., Gates, P. & Harris, N. 'Pod shatter' in *Arabidopsis thaliana*, *Brassica napus* and *B. juncea*. *J. Microsc.* **181**, 195–203 (1996).
- Bowman, J. L., Baum, S. F., Eshed, Y., Putterill, J. & Alvarez, J. Molecular genetics of gynoecium development in *Arabidopsis*. *Curr. Top. Dev. Biol.* **45**, 155–205 (1999).

- Ferrández, C., Pelaz, S. & Yanofsky, M. Control of carpel and fruit development in *Arabidopsis*. *Annu. Rev. Biochem.* **68**, 321–354 (1999).
- Rollins, R. C. *The Cruciferae of Continental North America—Systematics of the Mustard Family from the Arctic to Panama* (Stanford Univ. Press, Stanford, 1993).
- Meakin, P. & Roberts, J. Dehiscence of fruit in oilseed rape (*Brassica napus* L.). I. Anatomy of pod dehiscence. *J. Exp. Bot.* **41**, 995–1002 (1990).
- Kempin, S. A. *et al.* Targeted disruption in *Arabidopsis*. *Nature* **389**, 802–803 (1997).
- Sundaresan, V. *et al.* Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**, 1797–1810 (1995).
- Eshed, Y., Baum, S. F. & Bowman, J. L. Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell* **99**, 199–209 (1999).
- Yanofsky, M. F. *et al.* The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* **346**, 35–40 (1990).
- Rounsley, S. D., Ditta, G. S. & Yanofsky, M. F. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* **7**, 1259–1269 (1995).
- Bowman, J. L., Drews, G. N. & Meyerowitz, E. M. Expression of the *Arabidopsis* floral homeotic gene *AGAMOUS* is restricted to specific cell types late in flower development. *Plant Cell* **3**, 749–758 (1991).
- Benfey, P. N. & Chua, N.-H. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**, 959–966 (1990).
- Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71**, 119–131 (1992).
- Gu, Q., Ferrández, C., Yanofsky, M. & Martienssen, R. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* **125**, 1509–1517 (1998).
- Mandel, A. M. & Yanofsky, M. F. The *Arabidopsis* *AGL8* MADS-box gene is expressed in inflorescence meristems and is negatively regulated by *APETALA1*. *Plant Cell* **7**, 1763–1771 (1995).
- Jack, T., Fox, G. L. & Meyerowitz, E. M. *Arabidopsis* homeotic gene *APETALA3* ectopic expression: transcriptional and post-transcriptional regulation determine floral organ identity. *Cell* **76**, 703–716 (1994).
- Blázquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).

Acknowledgements

We thank C. Ferrández, A. Sessions, S. Kempin, A. Pinyopich, E. Alvarez-Buylla, J. Spence and N. Harris for helpful discussions; K. Feldmann and his lab for the gift of DNA and seeds from his T-DNA insertional collection; R. Martienssen for providing GT140 seed; and S. Guimil, T. Khammungskhune, C. Chien, H. Cartwright and A. Roeder for assistance with the *shp1* and *shp2* mutagenesis screens. This work was supported by grants from the National Science Foundation, the National Institutes of Health, Monsanto Company and the University of California BioSTAR programme.

Correspondence and requests for materials should be addressed to M.Y. (e-mail: marty@ucsd.edu).

Rapid degradation of a large fraction of newly synthesized proteins by proteasomes

Ulrich Schubert*†, Luis C. Antón*, James Gibbs*, Christopher C. Norbury*, Jonathan W. Yewdell* & Jack R. Bennink*

*Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA

†Heinrich-Pette Institute, University of Hamburg, Hamburg, Germany

MHC class I molecules function to present peptides eight to ten residues long to the immune system. These peptides originate primarily from a cytosolic pool of proteins through the actions of proteasomes¹, and are transported into the endoplasmic reticulum, where they assemble with nascent class I molecules². Most peptides are generated from proteins that are apparently metabolically stable. To explain this, we previously proposed that peptides arise from proteasomal degradation of defective ribosomal products (DRiPs). DRiPs are polypeptides that never attain native structure owing to errors in translation or post-translational processes necessary for proper protein folding³. Here we show, first, that DRiPs constitute upwards of 30% of newly synthesized proteins as determined in a variety of cell types; second, that at least some DRiPs represent ubiquitinated proteins; and last, that ubiquitinated DRiPs are formed from human immunodeficiency virus Gag polyprotein, a long-lived viral protein that serves as a source of antigenic peptides.

Reasoning that proteasomes should be the principal consumer of DRiPs, we examined the effects of the proteasome inhibitors carbobenzoxy-LeuLeu-leucinal (zLLL) and lactacystin on the recovery of proteins from HeLa cells pulse-radiolabelled with [35 S]methionine ([35 S]Met) for 30 seconds and chased for up to 60 min. Lactacystin is a highly specific inhibitor of the proteasome⁴, whereas zLLL inhibits proteasomes and other proteases⁵. Radiolabelled cells were either directly boiled in SDS–polyacrylamide gel electrophoresis (PAGE) sample buffer (Fig. 1, 'All'), or lysed by freeze–thawing and fractionated by centrifugation into soluble cytosolic ('Cyto') and insoluble (pellet; 'Ins') fractions before boiling in sample buffer. Radiolabelled proteins were resolved by SDS–PAGE and the quantities of radioactivity in dried gels were measured. For all matched inhibitor-treated and untreated samples, identical amounts of cellular protein were loaded per lane. Greater cellular equivalents were loaded for fractionated than for unfractionated samples; 'All' lanes therefore contained less radioactivity than the sum of the 'Cyto' and 'Ins' lanes.

In the absence of proteasome inhibitors, the amount of protein recovered reached plateau values within 10 min following pulse radiolabelling (Fig. 1a, b). This lag presumably reflects the time needed to utilize [35 S]Met-charged tRNA. Incubation of cells with lactacystin/zLLL progressively increased the amount of protein recovered in the 'All' fraction, reaching threefold the amount recovered from untreated cells by 60 min. This cannot be attributed

to an increased rate of translation or [35 S]Met incorporation because first, all of the increase occurred in the insoluble fraction, and second, the increase was biased towards slower-migrating proteins. The simplest interpretation of these findings is that a large fraction (two-thirds) of proteins synthesized under these conditions are degraded during or rapidly after translation by proteases sensitive to lactacystin/zLLL. On the basis of additional evidence presented below, we believe that they consist largely of DRiPs and refer to them as such in what follows.

The consequences of preincubating cells for increasing durations with lactacystin/zLLL before radiolabelling were monitored by acid precipitation of extracts on filter paper as well as by SDS–PAGE (Fig. 1c). The enhancing effect of lactacystin/zLLL on protein recovery was maximal when the inhibitors were added 10 min before radiolabelling; preincubation for 30 min resulted in a slight (10%) decrease in the effect. Pretreating cells for 60 min had only a minor effect on protein recovery, whereas treating for 2 or 4 h profoundly decreased protein synthesis. SDS–PAGE analysis of insoluble fractions revealed that an increased recovery of proteins of relative molecular mass (M_r) 46,000–250,000 (46K–250K) was observed only for cells exposed to lactacystin/zLLL for 10 or 30 min before radiolabelling. In an additional experiment (not shown), exposure of cells to 50 μ M zLLL immediately after pulse radiolabelling did not enhance protein recovery or modify the size or distribution of proteins. Thus, the optimal detection of DRiPs

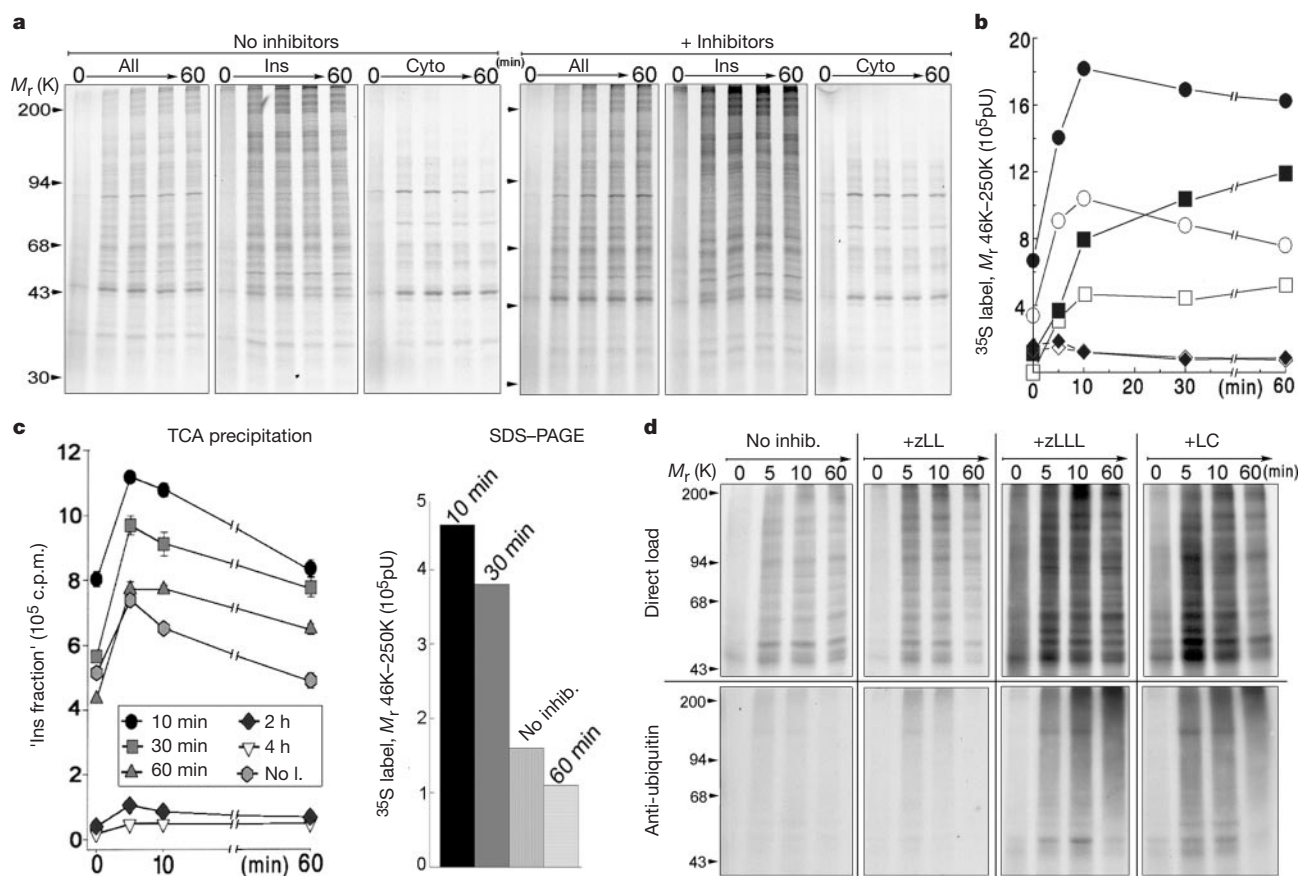


Figure 1 Detection of cellular DRiPs. HeLa cells treated with 50 μ M of zLLL/lactacystin each during the final 30 min of a 60-min starvation in Met-free media were radiolabelled for 30 s and chased for up to 60 min. **a**, Fluorographs of fraction samples separated by SDS–PAGE. All, total proteins; Ins, insoluble fraction; Cyto, cytoplasmic fraction. **b**, Quantification of the 35 S-labelled proteins in the M_r 46K–250K range plotted in PhosphorImager units (p.u.). Filled circles, insoluble fraction plus inhibitors; open circles, insoluble fraction without inhibitors; filled squares, total proteins plus inhibitors; open squares, total proteins without inhibitors; filled diamonds, cytoplasmic fraction plus inhibitors; open diamonds, cytoplasmic fraction without inhibitors. **c**, HeLa cells were treated with zLLL/lactacystin for the indicated durations and radiolabelled proteins recovered in the 'Ins' fraction by precipitation with TCA were quantified in triplicate. 'Ins' fractions from the 60-min chase samples were analysed by SDS–PAGE and quantified as in **a**. No inh., no inhibitors. **d**, HeLa cells were pulse-radiolabelled after incubation with 50 μ M zLLL, zLLL or lactacystin (LC) starting 15 min before the pulse. The 'Ins' fraction was analysed by SDS–PAGE directly or after extraction with SDS, renaturation and binding of material to ubiquitin-specific antibodies.

requires a narrow window of exposure time to proteasome inhibitors, just long enough, in fact, to block proteasomes before protein synthesis. In an additional experiment we examined the influence of Met starvation on DRiP formation by pulse radiolabelling for 2.5 min with or without prior incubation in Met-deficient medium. Prior Met starvation doubled the quantities of DRiPs detected, possibly owing to protein misfolding induced by limitations in Met-charged tRNA.

If DRiPs truly represent misfolded or unpartnered proteins, it is expected that at least a fraction of DRiPs are ubiquitinated. This was examined by immunoprecipitation of SDS extracts from fractionated [35 S]Met-labelled HeLa cells with ubiquitin-specific antibodies. Cells were treated with either 50 μ M zLLL, lactacystin or carbobenzoxy-Leu-leucinal (zLL), which like zLLL is a potent inhibitor of thiol proteases but does not detectably inactivate proteasomes at the concentration used⁶. Figure 1d shows the effects of the inhibitors on the recovery of proteins in the post-nuclear 150,000g pellet, where it is seen that zLLL and lactacystin have similar effects on protein recovery when used individually (as in

Fig. 1a–c). zLL has no significant effect on DRiP recovery, strongly implicating proteasomes in DRiP degradation. Quantification of immunoprecipitated material showed that lactacystin and zLLL increase by ~ 4.5 -fold the recovery of ubiquitinated material relative to untreated or zLL-treated cells. As after direct loading on gels, most anti-ubiquitin-reactive material migrated at $M_r > 43$ K.

We next examined whether we could detect DRiPs generated from a specific gene product, the Pr55 Gag polyprotein precursor of human immunodeficiency virus type 1 (HIV-1), which is the source of a large variety of class I antigenic peptides. HeLa cells were transfected with an infectious molecular clone of HIV-1, and the soluble material in high-detergent-concentration extracts from pulse-radiolabelled cells that bound to a mixture of antibodies raised against the major Pr55 processing product capsid (CA), was analysed by SDS–PAGE (Fig. 2a). In cells treated with zLLL/lactacystin just before labelling we detected, in addition to the expected Gag precursor protein Pr55, an increasing intensity of individual diffuse bands in addition to a smear of proteins between them migrating from M_r 60K to the top of the stacking gel. The magnitude of the increase (Fig. 2b) suggests that $\sim 40\%$ of the specific radioactivity recovered by anti-CA antibodies consists of high- M_r Gag-DRiPs. The smear recovered by anti-CA antibodies was not recovered under the same conditions from cells transfected with a control plasmid nor when non-transfected [35 S]Met-labelled cells were mixed with unlabelled HIV-1-expressing cells before extraction (not shown). This strongly suggests that the smear truly represents Gag-DRiPs and not cellular proteins that bind non-specifically to Gag–anti-Gag immune complexes. In the same experiment there was also a large increase in the amount of radiolabelled polyubiquitinated proteins that bound to a monoclonal antibody, FK2, specific for polyubiquitinated proteins⁷ (Fig. 2c). We next demonstrated that the high- M_r smear recovered with HIV-specific antibodies from cells treated with proteasome inhibitors represents polyubiquitinated viral proteins by showing that the material reacted in western blots with the FK2 monoclonal antibody (Fig. 2d). This was a specific effect, requiring that cells both express HIV proteins and be exposed to proteasome inhibitors. These findings show that ubiquitinated forms of a defined long-lived viral gene product are produced rapidly after synthesis and are destroyed by proteasomes.

If DRiPs are abundant and largely ubiquitinated, then blocking protein synthesis should rapidly deplete polyubiquitinated protein pools. This prediction was confirmed when HeLa cells were treated

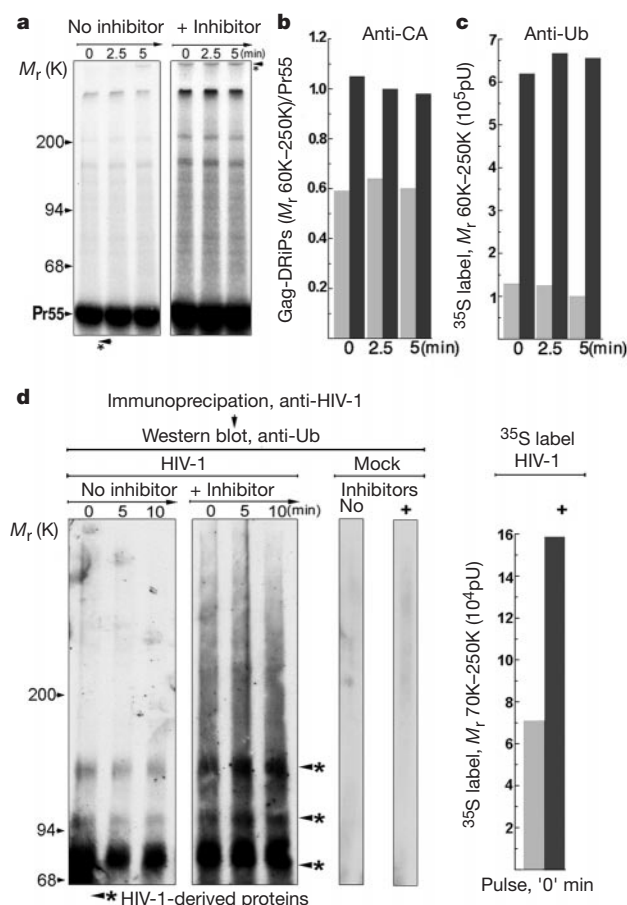


Figure 2 Detection of viral DRiPs. HeLa cells transfected with HIV-1 clone pNL4-3 were treated with 25 μ M lactacystin/zLLL before being radiolabelled as in Fig. 1. Material recovered with anti-CA antibodies was separated by SDS–PAGE (**a**) and quantified as the ratio of counts recovered in the M_r 60K–250K smear relative to counts recovered as Pr55 Gag precursor protein (**b**). **c**, Enhanced amounts of high- M_r polyubiquitinated proteins were recovered from the same lysates by using the FK2 monoclonal antibody. **d**, Lysates of a pulse–chase similar to that shown in **a** with pNL4-3 and mock-transfected HeLa cells were immunoprecipitated with anti-HIV antibodies, and polyubiquitinated proteins present in the immunoprecipitates were detected by western blotting with FK-1 monoclonal antibody. Major bands migrating in the M_r range 70K–160K labelled with an asterisk represent HIV-specific proteins that reacted with FK-2. Relative amounts of 35 S-labelled HIV-DRiPs migrating in the M_r range 70K–250K were quantified for the pulse time point (right panel). In **b**, **c** and **d**, grey columns, no inhibitors; black columns, with inhibitors.

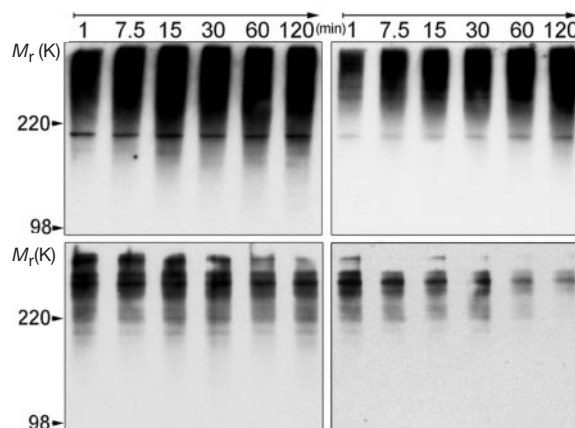


Figure 3 Dependence of polyubiquitinated proteins on continuing protein synthesis. HeLa cells were treated with or without protein synthesis inhibitors in the presence or absence of 25 μ M zLLL for up to 120 min. The relative levels of polyubiquitinated proteins were detected by western blotting of total cell lysates with the FK-2 monoclonal antibody specific for polyubiquitin. Left panels, with protein synthesis; lower panels, without protein synthesis; top panels, with zLLL; lower panels, without zLLL.

with protein synthesis inhibitors in the presence or absence of proteasome inhibitors, and levels of polyubiquitinated proteins were determined in western blots with the FK2 monoclonal antibody. In zLLL-treated cells, blocking protein synthesis retarded the accumulation of polyubiquitinated proteins (Fig. 3). This echoes earlier findings that protein synthesis is required for the accumulation of ubiquitin at the microtubule-organizing centre⁸, now known to be a specialized location for protein degradation^{9,10}. Importantly, in the absence of proteasome inhibitors, blocking protein synthesis rapidly depletes polyubiquitinated proteins, providing evidence for the existence of DRiPs in cells not exposed to proteasome inhibitors. These data suggest that DRiPs are a significant source of ubiquitinated proteins and that a substantial proportion of DRiPs are polyubiquitinated. Given the high level of DRiP ubiquitination, it was

possible that a significant fraction of the signal detected in the high- M_r ^{35}S -labelled DRiPs emanated from Met residues in ubiquitin synthesized *de novo*. This possibility was eliminated by detecting DRiPs after labelling with [^{35}S]Cys, a residue not present in ubiquitin (data not shown).

The experiments described so far used HeLa cells, which are highly aneuploid. We extended these results to dendritic cells, which are believed to function *in vivo* to present endogenously synthesized viral antigens to naive T cells¹¹. With the use of a dendritic cell line derived from mouse bone marrow, DRiPs also constituted a significant fraction of newly synthesized proteins (see Supplementary Information). Although it is not technically feasible to label cells *in vivo*, we approached this ideal by also measuring DRiPs in lymph node cells as quickly as possible after their removal from mice (Fig. 4a). The results completely recapitulated the findings made with dendritic and HeLa cells, with the added feature that, after preservation of the stacking gel in this experiment, it was clear that DRiPs were also present as high- M_r material that barely entered the stacking gel. We estimate that, under these conditions, DRiPs constitute 30% of newly synthesized proteins on the basis of the differences in total radioactivity detected for each time point in fluorograms of Fig. 4a.

We tested the prediction that DRiPs are a significant source of peptide ligands for class I molecules by examining the effect of blocking protein synthesis on the transport of mouse class I molecules H-2D^b and H-2K^b class I molecules from the endoplasmic reticulum, which is known to be controlled by the availability of suitable peptide ligands¹². The export of nascent class I molecules was assessed by collecting class I molecules pulse-radiolabelled with [^{35}S]Met and tracking the decreased electrophoretic mobility associated with the sialylation of *N*-linked oligosaccharides in the *trans*-Golgi complex (Fig. 4b). As reported previously¹³, K^b is exported more rapidly than D^b. Notably, the export of both allomorphs is slowed 2–3-fold by blocking protein synthesis immediately after labelling. In contrast, the export of the transferrin receptor was unaffected, showing that the effect on class I molecules is not due to general retardation in membrane protein export from the endoplasmic reticulum.

We believe that this is the first study to examine the overall efficiency of protein biogenesis. Using the most physiological cell system (lymph node cells *ex vivo*) under the most physiological conditions (no Met starvation), one-third of newly synthesized proteins are rapidly destroyed by proteasomes. We note that the precise proportion of DRiPs relative to truly short-lived proteins, as well as the relative contributions of errors in mRNA synthesis, processing and translation versus errors in protein folding remain to be directly established. But these findings, in conjunction with others¹⁴, firmly link the generation of class I peptide ligands to continuing protein synthesis. The fact that self peptides recovered from class I molecules in sufficient quantities for chemical identification demonstrate no apparent molar bias towards derivation from short-lived proteins^{15,16} indicates strongly that most of the rapidly degraded proteins that we identify in the present study are in fact DRiPs derived mostly from long-lived proteins. The 70% efficiency for protein biosynthesis compares favourably to an estimated 50% efficiency for producing functional mRNA from heterogeneous nuclear RNA¹⁷. The remarkably high abundance of proteasomes (constituting 1% of total cellular protein) is consistent with a high constitutive level of DRiPs requiring disposal. We expect that some proteins are more disposed than others towards becoming DRiPs, on the basis of size and inherent difficulties in folding or assembly. If DRiPs maintain some elements of native structure, they might act in a dominant-negative manner to interfere with the function of proteins that associate with normally folded domains of DRiPs, as seems to occur with HIV-1 Gag (U.S., unpublished data). If under some conditions DRiP degradation were delayed, this could contribute to disease processes. □

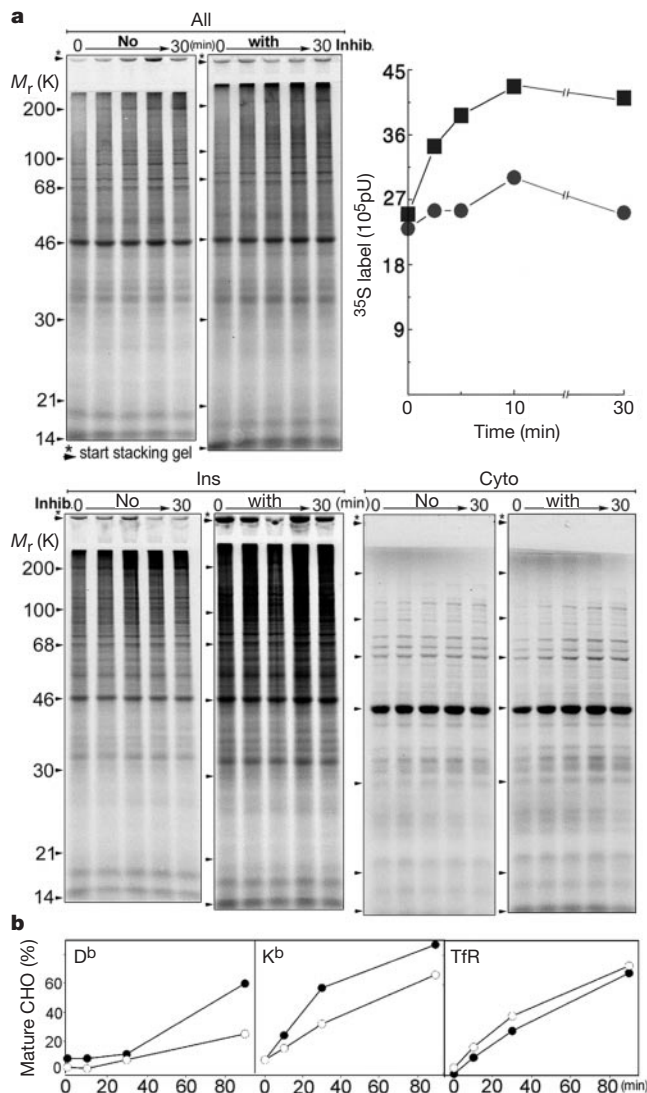


Figure 4 DRiP formation in lymph node cells and effect of protein synthesis inhibitors on class I transport. **a**, Lymph node cells were radiolabelled for 2.5 min in the absence or presence of 10 μM zLLL, 10 μM lactacystin and 1 μM clasto-lactacystin β -lactone. Total ('All') or fractionated ('Ins' and 'Cyto') cell lysates were separated by SDS-PAGE; in the graph the relative amounts of proteins in the M_r range 14K–250K detected in the 'All' fraction are plotted against the duration of the chase. Circles, no inhibitors; squares, with inhibitors. **b**, Pulse-radiolabelled RMA cells were chased for the indicated durations in the presence (open circles) and absence (filled circles) of protein synthesis inhibitors, and K^b, D^b and transferrin receptor (Tfr) molecules were recovered with specific monoclonal antibodies. Each class I specific monoclonal antibody was used twice sequentially to insure quantitative recovery. Shown is a time-dependent shift in mobility in SDS-PAGE quantified by PhosphorImager analysis. CHO, *N*-linked carbohydrate.

Methods

The cultivation of RMA and HeLa cell lines, as well as the generation of dendritic and lymph node cells are described in detail elsewhere (see Supplementary Information). Different cells vary considerably in their sensitivities to the various proteasome inhibitors, in terms of efficacy and adverse effects (decreased protein synthesis and viability). The amounts and concentrations of inhibitors used for different cells were optimized. For pulse–chase experiments, cells were incubated at 37 °C for up to 60 min in Met-free, serum-free RPMI (M⁺RPMI) before being radiolabelled. Cells (usually ~10⁷) were pelleted, resuspended in 150 µl M⁺RPMI prewarmed to 37 °C and radiolabelled by adding prewarmed [³⁵S]methionine (Amersham Life Science) to a final concentration of 5 mCi ml⁻¹. After incubation of cells at 37 °C for 30 or 150 s, 1 ml ice-cold Iscove's DMEM with 10% fetal bovine serum (I⁺) and supplemented with 10 mM Met (M⁺I⁺) was added, and cells were microcentrifuged for 10 s at 8,000g, aliquoted into 1 ml prewarmed M⁺I⁺, and incubated on a rotating platform at 37 °C during the chase period. In experiments in which cells were pulse-labelled for 30 s, cells were washed and aliquoted in ice-cold M⁺I⁺. Cells were lysed by 3 cycles of freeze–thawing (37 °C water bath alternated with solid CO₂) in isotonic buffer (0.25 M sucrose, 10 mM triethanolamine, 1 mM EDTA, 20 mM acetic acid pH 7.4) containing 1 mM phenylmethylsulphonyl fluoride (PMSF), 5 mM N-ethyl maleimide, 20 µM ZLL, 20 µM lactacystin, 100 units ml⁻¹ DNase I and Complete protease inhibitor cocktail (Boehringer Mannheim). Insoluble material was separated from the soluble cytosolic fraction by centrifugation of cell lysates at 150,000g for 2 h at 4 °C. For the experiment shown in Fig. 1d, nuclei were removed before ultracentrifugation by centrifugation at 3,000g for 10 min at 4 °C.

Aliquots of unfractionated cells resuspended in isotonic buffer before freeze–thawing ('All' fraction) were immediately mixed with boiling sample buffer (2% (w/v) SDS, 1% (v/v) 2-mercaptoethanol, 1% (v/v) glycerol, 65 mM Tris-HCl pH 6.8, 20 µM ZLL, 20 µM lactacystin). Pellets of insoluble material after ultracentrifugation were resuspended in Chaps-deoxycholate buffer (50 mM Tris-HCl pH 8.0, 5 mM EDTA, 100 mM NaCl, 0.5% (w/v) Chaps, 0.2% (w/v) deoxycholate) containing inhibitors as above, and equivalent amounts of soluble and insoluble samples were injected into boiling sample buffer. Samples were separated in 12.5% (w/v) Acryl aide Prosieve gels (FMC Bioproducts). Gels were fixed for 30 min in 40% (v/v) methanol, 10% (v/v) acetic acid, rinsed with water, soaked in 1 M sodium salicylic acid for 3 h, and dried. Radioactivity in gels was quantified with a PhosphorImager (Molecular Dynamics). For precipitations with trichloroacetic acid (TCA), aliquots of cell lysates were spotted onto DEAE-cellulose, washed with 10% (w/v) TCA and twice more with 70% ethanol then dried; radioactivity was then determined with a β-plate counter (MicroBeta 1450 Trilux; Wallac Oy). All estimations were done in triplicate.

Immunoprecipitation with ubiquitin-specific antibodies (mixture of polyclonal rabbit IgGs from several commercial sources) prebound to Protein G–Sepharose was performed on protein samples that had been denatured in SDS under non-reducing conditions, precipitated with acetone, resuspended in Chaps-deoxycholate buffer containing inhibitors as above, and renatured in Triton buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.4, 0.1% (v/v) Triton X-100, 1 mM PMSF). For western blot analysis, HeLa cells were cultured in six-well plates and treated with a mixture of protein synthesis inhibitors (25 µg ml⁻¹ cycloheximide, 25 µg ml⁻¹ emetine and 25 µg ml⁻¹ puromycin) or proteasome inhibitors. Cells were lysed while attached on the plate with Chaps-deoxycholate buffer containing inhibitors as above. Aliquots of lysates were separated in 8% Prosieve gels, electrotransferred to Immobilon P membranes (Millipore), probed with mouse monoclonal antibody specific for polyubiquitin (clone FK-1; Nippon Bio-Test Laboratories), and developed by using the ECL system (Pierce).

For detection of HIV-1 Gag-DRiPs HeLa cells were transfected with pNL4-3 (ref. 18) by using LipofectAMINE 2000 (Gibco BRL) and a standard pulse–chase was conducted 24 h later. After the chase, cells were frozen on solid CO₂, thawed and instantly lysed in Chaps buffer containing 1% deoxycholate and protease inhibitors as above. After being precleared with preimmune serum, aliquots of lysates were immunoprecipitated with γ-globulin from pooled plasma of HIV-1 donors (NIH AIDS Research and Reference Reagent Program, catalogue number 192) and a cocktail of six monoclonal and three polyclonal antibodies directed against various epitopes of HIV-1 CA derived from native, recombinant and synthetic HIV-1 Gag proteins (catalogue numbers 287, 384, 1238, 1239, 1241, 1244, 4121 and 4250), as well as against full-length recombinant CA produced in *Escherichia coli* (Seramun GmbH). Polyubiquitinated forms of anti-HIV-1 immunoprecipitates separated in 6% Acryl aide gels were transferred to membranes and probed with the FK-1 monoclonal antibody. The pulse–chase experiment and the analysis of class I transport in RMA cells are described in detail elsewhere (see Supplementary Information).

Received 3 December 1999; accepted 9 February 2000.

1. Rock, K. L. & Goldberg, A. L. Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **17**, 739–779 (1999).
2. Pamer, E. & Cresswell, P. Mechanisms of MHC class I restricted antigen processing. *Annu. Rev. Immunol.* **16**, 323–358 (1998).
3. Yewdell, J. W., Antón, L. C. & Bennink, J. R. Defective ribosomal products (DRiPs). A major source of antigenic peptides for MHC class I molecules? *J. Immunol.* **157**, 1823–1826 (1996).
4. Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268**, 726–731 (1995).
5. Rock, K. L. *et al.* Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761–771 (1994).
6. Vinititsky, A., Michaud, C., Powers, J. C. & Orłowski, M. Inhibition of the chymotrypsin-like activity of the pituitary multicatalytic proteinase complex. *Biochemistry* **31**, 9421–9428 (1992).
7. Fujimuro, M., Sadada, H. & Yokosawa, H. Production and characterization of monoclonal antibodies specific to multi-ubiquitin chains of polyubiquitinated proteins. *FEBS Lett.* **349**, 173–180 (1994).

8. Wojcik, C., Schroeter, D., Wilk, S., Lamprecht, J. & Pawletz, N. Ubiquitin-mediated proteolysis centers in HeLa cells: indication from studies of an inhibitor of the chymotrypsin-like activity of the proteasome. *Eur. J. Cell Biol.* **71**, 311–318 (1996).
9. Johnston, J. A., Ward, C. L. & Kopito, R. Aggresomes: a cellular response to misfolded proteins. *J. Cell Biol.* **143**, 1883–1898 (1998).
10. Anton, L. C. *et al.* Intracellular localization of proteasomal degradation of a viral antigen. *J. Cell Biol.* **146**, 113–124 (1999).
11. Steinman, R. M. in *Fundamental Immunology* (ed. Paul, W. E.) 547–604 (Lippincott–Raven, Philadelphia, 1998).
12. Townsend, A. *et al.* Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* **340**, 443–448 (1989).
13. Degen, E. & Williams, D. B. Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility molecules. *J. Cell Biol.* **112**, 1099–1115 (1991).
14. Reits, E. A. J., Vos, J. C., Grommé, M. & Neefjes, J. The major substrates for TAP *in vivo* are derived from newly synthesized proteins. *Nature* **404**, 774–778 (2000).
15. Rammensee, H. -G., Bachmann, J. & Stevanovic, S. *MHC Ligands and Peptide Motifs* (Landes Bioscience, Austin, 1997).
16. Engelhard, V. H. Structures of peptides associated with class I and class II MHC molecules. *Annu. Rev. Immunol.* **12**, 181–207 (1994).
17. Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland, New York, 1994).
18. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* **59**, 284–291 (1986).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Buschling for technical assistance. C.N. is the recipient of a Wellcome Prize Traveling Fellowship; U.S. was supported by grant Schu11/2-1 and a Heisenberg grant from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to J.W.Y. (e-mail: jyewdell@nih.gov) or J.R.B. (e-mail: jbenink@nih.gov).

The major substrates for TAP *in vivo* are derived from newly synthesized proteins

Eric A. J. Reits, Jan C. Vos, Monique Grommé & Jacques Neefjes

Division of Tumor Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

The transporter associated with antigen processing (TAP) is a member of the family of ABC transporters that translocate a large variety of substrates across membranes¹. TAP transports peptides from the cytosol into the endoplasmic reticulum for binding to MHC class I molecules and for subsequent presentation to the immune system². Here we follow the lateral mobility of TAP in living cells. TAP's mobility increases when it is inactive and decreases when it translocates peptides. Because TAP activity is dependent on substrate, the mobility of TAP is used to monitor the intracellular peptide content *in vivo*. Comparison of the diffusion rates in peptide-free and peptide-saturated cells indicates that normally about one-third of all TAP molecules actively translocate peptides. However, during an acute influenza infection TAP becomes fully employed owing to the production and degradation of viral proteins. Furthermore, TAP activity depends on continuing protein translation. This implies that MHC class I molecules mainly sample peptides that originate from newly synthesized proteins, to ensure rapid presentation to the immune system.

TAP is composed of two subunits, TAP1 and TAP2, that form a structure consisting of a transmembrane domain, a peptide-binding domain and two nucleotide-binding domains^{3,4}. To investigate whether conformational changes in TAP would affect its lateral

mobility, we stably transfected the human melanoma cell line Mel JuSo⁵ with the human TAP1 subunit tagged with green fluorescent protein (GFP) at its cytosolic carboxy terminus. Positive clones were selected for low fluorescence to ensure optimal assembly of TAP1–GFP into TAP1–TAP2 complexes. These complexes show typical staining of endoplasmic reticulum with fluorescence in the nuclear envelope and the endoplasmic reticulum network (Fig. 1a). To determine the lateral diffusion of TAP–GFP, we used fluorescence recovery after photobleaching (FRAP) techniques. A spot in the endoplasmic reticulum is bleached and the recovery of fluorescence over time is measured, which is due to the diffusion of surrounding TAP–GFP molecules into the bleached spot. Assuming that fluorescence recovery follows a two-dimensional model of diffusion, the diffusion coefficient D is inversely proportional to the fluorescence recovery half-time $t_{1/2}$ (see Methods). For Mel JuSo cells cultured at 37°C under standard conditions⁵, the diffusion coefficient of TAP–GFP is $(3.92 \pm 0.16) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$ (Fig. 1b, column C). This is comparable with values found for other membrane proteins^{6–8}. To increase the rate of peptide translocation by TAP, we microinjected cells with a nonamer substrate peptide p417, which does not bind to the MHC class I molecules expressed in Mel JuSo⁹. As shown in Fig. 1b, TAP mobility is decreased until saturated levels of peptide are reached ($D = (3.18 \pm 0.29) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$). Upon saturation, all TAP complexes are likely to be active. In contrast, TAP mobility increases ($D = (4.45 \pm 0.29) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$) when cells are depleted of endogenous peptide substrates by inhibiting the proteasome with lactacystin¹⁰. This confirms the importance of the proteasome in peptide generation. The same increase in mobility is reached when cells are depleted of ATP, which

is necessary for peptide translocation^{9,11}. Taken together, these results show that peptide translocation by TAP results in a reduced lateral mobility.

Next, we microinjected a peptide with a very long side chain into cells, which binds to TAP but cannot be translocated owing to its size¹². TAP mobility is considerably lower than observed for conventional peptides (Fig. 1c; $D = (2.6 \pm 0.23) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$), which is lactacystin-independent. This effect requires ATP and suggests that decreased TAP mobility is caused by a peptide- and ATP-dependent conformational change in the TAP complex. This is independent of peptide translocation into the endoplasmic reticulum lumen and the long-side-chain peptide might trap TAP in an 'open' conformation. The same peptide, but with a shorter side chain that can be translocated into the endoplasmic reticulum by TAP¹², behaves like peptide p417 (data not shown).

Before peptide loading, many empty class I molecules such as HLA-A2 and HLA-A2–GFP^{13,14} efficiently associate with TAP and form a large protein complex incorporating tapasin and the endoplasmic reticulum chaperones calreticulin and Erp57 (ref. 15). To determine whether class I association or dissociation influences TAP mobility during our experiments, we performed FRAP experiments on the endoplasmic reticulum of Mel JuSo cells stably transfected with GFP-tagged HLA-A2. HLA-A2–GFP has a lateral mobility comparable to that of TAP (Fig. 1d; $D = (3.87 \pm 0.22) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$), as was previously shown for the murine class I allele H2-L^d–GFP (ref. 8). Class I mobility is reduced in an ATP-dependent fashion upon microinjection of long-side-chain-containing peptides that do not enter the endoplasmic reticulum¹². This implies that there is a stable association of TAP with HLA-A2–GFP, and that

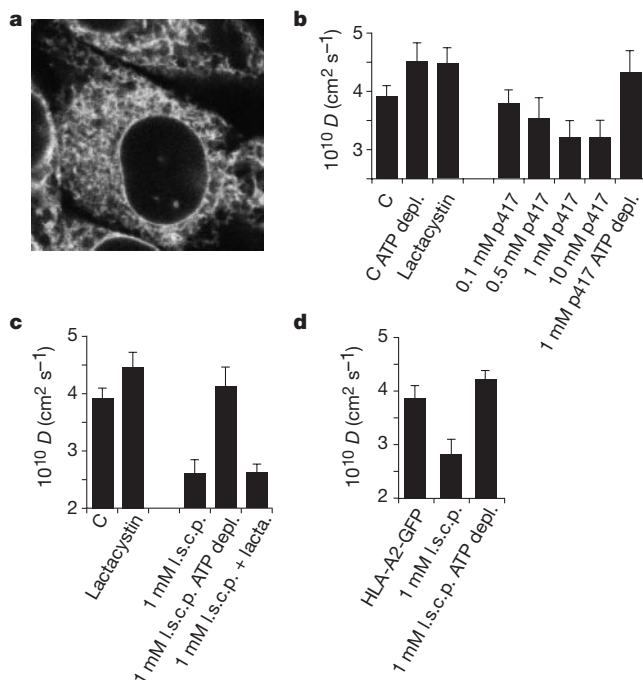


Figure 1 Analysis of TAP activity *in vivo* as measured by lateral mobility. **a**, TAP1–GFP distribution in living Mel JuSo cells (scale bar 2 μm). **b**, Lateral mobility of TAP in Mel JuSo cells in the presence or absence of ATP, lactacystin or various concentrations of peptides (in the microinjection needle). C, non-manipulated cells. **c**, Lateral mobility of TAP in Mel JuSo cells microinjected with the non-translocating long-side-chain-containing peptide (l.s.c.p.) in the presence or absence of ATP or lactacystin. **d**, Lateral mobility of GFP-tagged HLA-A2 in the endoplasmic reticulum of Mel JuSo cells microinjected with l.s.c.p.

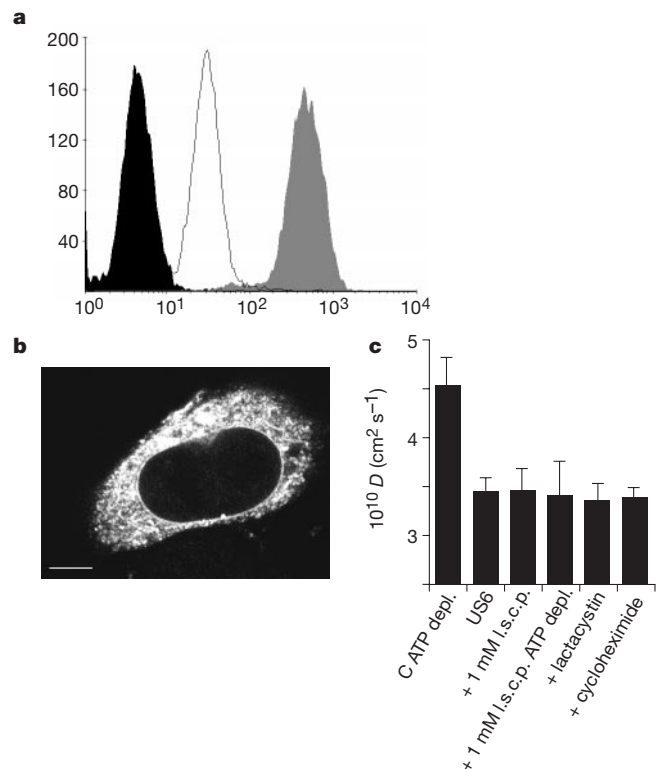


Figure 2 Soluble US6 blocks mobility changes of TAP. **a**, FACS analysis of cell surface expressed class I molecules on cells transfected with TAP1–GFP alone (grey peak) or together with soluble US6 (broken line). As a control, cells were incubated with the second antibody only (black peak). **b**, Soluble US6 distribution in stably transfected Mel JuSo cells (scale bar 2 μm). **c**, Lateral mobility of TAP1–GFP in Mel JuSo cells stably transfected with both TAP1–GFP and soluble US6 in the presence or absence of microinjected long-side-chain-containing peptide (l.s.c.p.), ATP or the proteasome inhibitor lactacystin.

non-class-I binding peptides alter the mobility of the TAP-MHC class I complex without entering the endoplasmic reticulum. In contrast to TAP-associated MHC class I molecules, GFP-tagged MHC class II molecules⁵ retained in the endoplasmic reticulum by treatment with brefeldin A for 2 hours diffuse twice as rapidly in the endoplasmic reticulum ($D = (6.5 \pm 1.4) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$). Class II mobility does not change upon the microinjection of peptide p417 ($D = (6.5 \pm 1.3) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$).

To investigate further the molecular basis for the peptide-induced changes in lateral mobility, we introduced the human cytomegalovirus protein US6 into TAP1-GFP cells. US6 interacts with TAP inside the endoplasmic reticulum lumen, and prevents peptide transfer but not the binding of peptide to TAP^{16,17}. GFP-tagged TAP-transfected Mel JuSo cells were co-transfected with the endoplasmic-reticulum-luminal domain of US6, and positive clones were selected for reduced MHC class I expression by fluorescence-assisted cell sorting (FACS; Fig. 2a). Expression of US6 in the endoplasmic reticulum was confirmed by immunohistochemistry (Fig. 2b) and western blotting (data not shown). When lateral mobility is measured, no changes are observed when peptides are introduced, upon depletion of ATP or treatment with lactacystin. This suggests that US6 blocks substrate- and ATP-dependent conformational changes of TAP from the endoplasmic reticulum luminal side (Fig. 2c; $D = (3.45 \pm 0.13) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$). The measured mobility of the TAP-US6 complex is lower than that of 'inactive' TAP in ATP- or peptide-depleted cells without US6. This is not due to the mass of the luminal domain of US6 (ref. 18), but probably reflects a conformational change of TAP induced by US6.

In contrast to Mel JuSo, the small lung carcinoma cell line GLC2 (ref. 19) expresses neither MHC class I nor the associated components tapasin and TAP (Fig. 3a). We reconstituted GLC2 with TAP1-GFP and TAP2 to determine whether the peptide-induced reduction in mobility is due to TAP alone or to alterations in the

TAP-class I loading complex. GLC2 cells transfected with TAP1-GFP and TAP2 show TAP distribution restricted to the endoplasmic reticulum (not shown), but neither tapasin nor class I are detected by western blotting (Fig. 3a). With the use of a peptide translocation assay with GLC2 transfectants, only GLC2 expressing both TAP1-GFP and TAP2 are able to translocate peptides into the endoplasmic reticulum in the presence of ATP (Fig. 3b). In contrast to GLC2 cells expressing TAP1-GFP only, GLC2 cells with both TAP subunits show the characteristic reduction in lateral mobility of TAP upon peptide microinjection (Fig. 3c). Together, these data suggest that alterations in the mobility of TAP are mainly the result of conformational changes related to pore opening (inhibited by US6) and peptide translocation (peptide- and ATP-dependent). The size of the TAP-class I loading complex⁸ has been estimated by using mathematics for the diffusion of complexes in uniform lipid membranes²⁰. Diffusion is related to temperature, viscosity and the radius of the protein complex in the lipid bilayer. Conformational changes in TAP might change the radius of the complex or they might result in a rotation and tilting of the transmembrane segments in the lipid bilayer²¹. The latter could affect the lipid-protein interface and thereby the local viscosity. Our experiments cannot distinguish between these possibilities. However, the fact that long-side-chain peptides reduce the mobility of TAP in GLC2 cells in the absence of associated proteins such as tapasin and MHC class I implies that conformational changes within TAP can markedly affect mobility.

We can now use alterations in the lateral mobility of TAP to

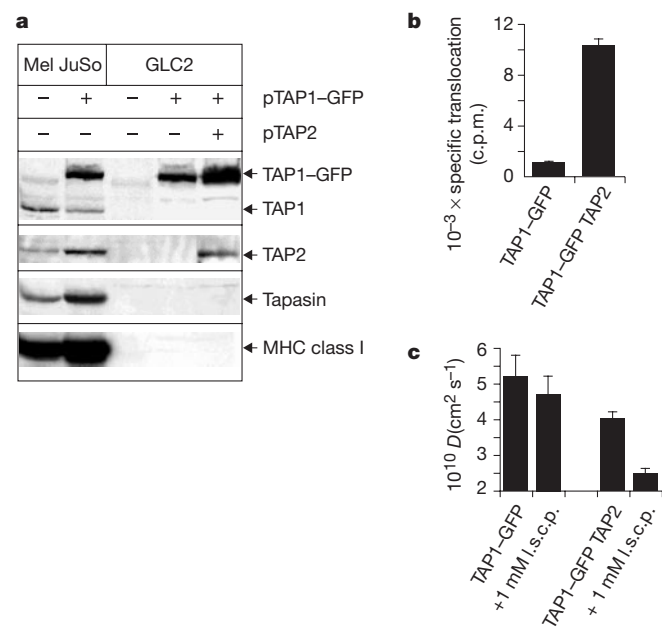


Figure 3 Peptide-induced changes in TAP expressed in GLC2 cells lacking tapasin and MHC class I molecules. **a**, Western blot analysis of total lysates of GLC2 transfected with the indicated constructs. The levels of TAP in Mel JuSo cells are shown for comparison. The proteins detected are indicated. **b**, ATP-dependent translocation of iodinated peptide p417 in microsomes of GLC2 transfected with TAP1-GFP either alone or in combination with TAP2. Values obtained in the absence of ATP were subtracted to determine specific translocation. **c**, Lateral mobility of TAP after the injection of long-side-chain-containing peptide (l.s.c.p.) in GLC2 cells transfected with TAP1-GFP either alone or in combination with TAP2.

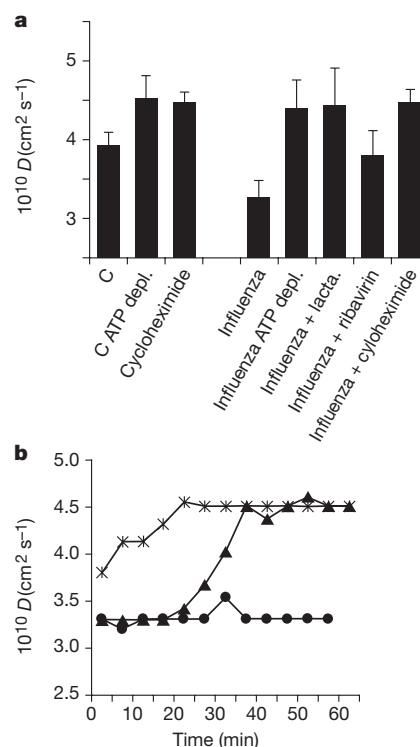


Figure 4 Newly synthesized proteins are the major source for peptides translocated by TAP. **a**, Lateral mobility of TAP 2 h after infection with influenza. Cells were incubated with the indicated inhibitors 1.5 h after infection. The inhibitory effect of treatment with ribavirin and cycloheximide on viral protein expression was confirmed by immunohistochemistry (data not shown). **b**, Representative FRAP time course following the availability of TAP substrate after the inhibition of protein synthesis. Uninfected (control, stars) and influenza-infected cells (triangles) were treated with cycloheximide at $t = 0$ and the lateral mobility of TAP1-GFP was determined every 5 min. As a control, cells were microinjected with saturated levels of peptide p417 (10 mM in the microinjection needle) and treated with cycloheximide (circles).

monitor the intracellular peptide content that is available to TAP as a substrate *in vivo*. Endogenous peptides supply around one-third of all TAP molecules in Mel JuSo cells with substrate, when the mobility of TAP in untreated cells is compared with lactacystin-treated cells and peptide-saturated cells (Fig. 1b). We tested whether the window of inactive TAP narrows after an acute viral infection. Infection of TAP-GFP Mel JuSo cells with influenza A virus results in a decreased TAP mobility within 2 hours, which is comparable to saturated levels of microinjected peptides (Fig. 4a; $D = (3.26 \pm 0.22) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$). This suggests that influenza infection rapidly increases the intracellular peptide pool available to TAP. The decrease in TAP mobility is both dependent on ATP and sensitive to proteasome inhibitors (Fig. 4a). To determine whether the increased peptide levels are derived from viral or host proteins, we used two inhibitors: ribavirin, which exclusively inhibits influenza RNA synthesis²², and cycloheximide, which blocks translational processes of both influenza and endogenous proteins. When virus-infected and control cells are treated for 30 minutes with cycloheximide, TAP mobility is increased to a similar level to that in lactacystin-treated or ATP-depleted cells ($D = (4.47 \pm 0.15) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$). Treatment of infected cells with ribavirin also increases TAP mobility ($D = (3.8 \pm 0.32) \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$), which indicates that the increased level of peptides in infected cells is derived mostly from viral proteins. In comparison with cycloheximide, ribavirin has less effect on TAP mobility. This is probably due to continuing host protein degradation and to the presence of stable viral mRNA synthesized before treatment with ribavirin, which will continue to be translated into viral proteins²³.

Surprisingly, the intracellular peptide pool was largely diminished by only 30 minutes of translation inhibition. To determine the turnover of proteins for the generation of TAP substrates more accurately, we measured changes in TAP mobility over time (Fig. 4b). When protein synthesis is blocked in non-infected cells, TAP activity increases within 15 minutes to levels comparable to those under ATP depletion. TAP mobility increases more slowly in influenza-infected cells treated with cycloheximide, probably owing to higher peptide levels at the start of the inhibition of protein synthesis. This indicates that the main substrates for TAP are derived from newly synthesized proteins. Microinjection of peptides in cycloheximide-treated cells excludes the possibility that cycloheximide inhibits peptide translocation by TAP. Whereas small quantities of peptides show time-dependent recovery of TAP mobility in both the presence and the absence of cycloheximide (data not shown), the mobility with saturating levels of peptide remains slow (Fig. 4b). In addition, cycloheximide has no effect on TAP mobility in the presence of US6 (Fig. 2b). These results suggest that most peptides destined for TAP are derived from newly synthesized proteins.

In a first step towards 'single cell biochemistry' we have shown that TAP mobility reflects TAP activity, with TAP molecules (and associating class I) moving more slowly when engaged in peptide translocation and faster when unoccupied. It is likely that a similar mobility-activity relationship will be observed for other ABC transporters as well. This observation is used to detect alterations in the intracellular peptide content, which rapidly increases after an acute viral infection owing to proteasome-generated viral peptides. Surprisingly, it seems that the pool of endogenous and viral peptides that constitute a substrate for TAP is mostly derived from newly synthesized proteins. Degradation of a relatively large pool of proteins shortly after synthesis might be the result of an apparently high failure rate of protein folding, as also suggested by Schubert *et al.*²⁴. As a consequence, MHC class I molecules are immediately loaded with antigenic fragments generated directly after protein translation without having to wait until the end of an antigen's properly folded life. Class I molecules will therefore be rapidly exposed to viral peptides, ensuring swift presentation to the surveying immune system. □

Methods

Constructs and antibodies.

TAP1-GFP was made by cloning full-length TAP1 into the vector pEGFP (Clontech). Soluble US6, residues 1–143 plus an additional 5 His codons, was generated by PCR (US6 template kindly provided by E. Wiertz). TAP2 was cloned upstream of an internal ribosomal entry site followed by an inactivated nerve growth factor receptor (IRES-ΔNGFR) fragment (provided by M. Heemskerk) to sort transfected cells for NGFR expression. Rabbit anti-US6 Ab (provided by E. Wiertz), mouse anti-TAP1 mAb and anti-TAP2 mAb (provided by R. Tampé), rabbit anti-tapasin (provided by P. Cresswell²⁵), mouse antibody against human class I complex (W6/32) and mouse mAb against human class I H chain (HC-10 (ref. 26)) were described previously. Long-side-chain peptide A(K-εK₂₀)DNATRDY¹², short chain peptide A(K-εK₈)DNATRDY¹² and p417 (ref. 9) were used in this study.

Confocal fluorescence microscopy analysis and FRAP experiments.

Confocal analysis of living and fixed cells was performed as described⁵ with a Leica TCS SP (Leica Microsystems) confocal system, equipped with an Ar-Kr laser. To inhibit proteasomes, cells were incubated at 37 °C in the presence of 10 μM lactacystin for 30 min. For ATP depletion, cells were incubated for 30 min with a mixture of sodium azide (0.05%) and 2-deoxyglucose (50 μM), conditions that blocked endocytosis²⁷. For influenza experiments, cells were incubated for 2 h with influenza A virus strain A/NT/60/68 at a concentration of 50 haemagglutination units per ml of medium (virus provided by G. Rimmelzwaan). This resulted in 95% infection efficiency, as verified by immunostaining with anti-haemagglutinin antibodies. Ribavirin (33 μg/ml) or cycloheximide (200 μM) was added 1.5 h after infection. For lateral diffusion measurements, a defined circular spot in the endoplasmic reticulum was bleached at full intensity, and an attenuated laser beam was used to monitor the recovery of fluorescence in the bleached region (with Leica TCS time-lapse photography). The size of the bleached spot was calculated by using fixed and subsequently bleached cells, resulting in a non-recovering, circular bleached spot with a radius of 0.8 μm. During FRAP experiments, other regions within the same cell and in an adjacent cell were monitored at the same time for loss of fluorescence due to diffusion and due to imaging, respectively (with Leica TCS physiology software). The half-time for recovery was calculated from each recovery curve after correction for the loss of fluorescence due to imaging (usually less than 4%). The diffusion coefficient, D , is described by the formula $D = w^2 \gamma_D / 4t_{1/2}$ (refs 28, 29). The invariable w denotes the half-width at e^{-2} times the height of the bleached spot (in our experiments w was 0.8 μm), and γ_D is dependent on the amount of bleaching and the profile of the laser intensity (in our experiment γ_D was 1.55). D was determined from at least three independent FRAP experiments (generally at least 10 cells per experiment) and depicted as means $\pm$ s.d. As a standard, FRAP experiments on non-manipulated cells were included as an internal control, and maintenance of a temperature of 37 °C in the culture chamber of the confocal laser scanning microscope was monitored. The calculated D , assuming two-dimensional diffusion in the endoplasmic reticulum, is probably an underestimate of the actual D owing to a time window between bleaching and the first recovery measurement of 0.6 s, and also the architecture of the endoplasmic reticulum.

Received 2 December 1999; accepted 10 February 2000.

- Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).
- Pamer, E. & Cresswell, P. Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* **16**, 323–358 (1998).
- Vos, J. C., Spee, P., Momburg, F. & Neefjes, J. Membrane topology and dimerization of the two subunits of the transporter associated with antigen processing reveal a three-domain structure. *J. Immunol.* **163**, 6679–6685 (1999).
- Vos, J. C., Reits, E. A. J., Wojcik-Jacobs, E. & Neefjes, J. Head-head/tail-tail relative orientation of the pore-forming domains of the heterodimeric ABC transporter TAP. *Curr. Biol.* **10**, 1–7 (2000).
- Wubolts, R. *et al.* Direct vesicular transport of MHC class II molecules from lysosomal structures to the cell surface. *J. Cell Biol.* **135**, 611–622 (1996).
- Kwon, G., Axelrod, D. & Neubig, R. R. Lateral mobility of tetramethylrhodamine (TMR) labelled G protein alpha and beta gamma subunits in NG 108-15 cells. *Cell Signal.* **6**, 663–679 (1994).
- Schlessinger, J. *et al.* Lateral transport on cell membranes: mobility of concanavalin A receptors on myoblasts. *Proc. Natl Acad. Sci. USA* **73**, 2409–2413 (1976).
- Marguet, D. *et al.* Lateral diffusion of GFP-tagged H2Ld molecules and of GFP-TAP1 reports on the assembly and retention of these molecules in the endoplasmic reticulum. *Immunity* **11**, 231–240 (1999).
- Neefjes, J. J., Momburg, F. & Hammerling, G. J. Selective and ATP-dependent translocation of peptides by the MHC-encoded transporters. *Science* **261**, 769–771 (1993).
- Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268**, 726–731 (1995).
- van Endert, P. M. *et al.* A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity* **1**, 491–500 (1994).
- Grommé, M. *et al.* The rational design of TAP inhibitors using peptide substrate modifications and peptidomimetics. *Eur. J. Immunol.* **27**, 898–904 (1997).
- Grommé, M. *et al.* Recycling MHC class I molecules and endosomal peptide loading. *Proc. Natl Acad. Sci. USA* **96**, 10326–10331 (1999).
- Neisig, A., Wubolts, R., Zang, X., Melief, C. & Neefjes, J. Allele-specific differences in the interaction of MHC class I molecules with transporters associated with antigen processing. *J. Immunol.* **156**, 3196–3206 (1996).
- Lehner, P. J. & Trowsdale, J. Antigen presentation: coming out gracefully. *Curr. Biol.* **8**, R605–R608 (1998).
- Ahn, K. *et al.* The ER-luminal domain of the HCMV glycoprotein US6 inhibits peptide translocation by TAP. *Immunity* **6**, 613–621 (1997).
- Hengel, H. *et al.* A viral ER-resident glycoprotein inactivates the MHC-encoded peptide transporter. *Immunity* **6**, 623–632 (1997).

18. Livneh, E. *et al.* Large deletions in the cytoplasmic kinase domain of the epidermal growth factor receptor do not affect its lateral mobility. *J. Cell Biol.* **103**, 327–331 (1986).
19. de Leij, L. *et al.* Characterization of three new variant type cell lines derived from small cell carcinoma of the lung. *Cancer Res.* **45**, 6024–6033 (1985).
20. Saffman, P. G. & Delbruck, M. Brownian motion in biological membranes. *Proc. Natl Acad. Sci. USA* **72**, 3111–3113 (1975).
21. Cha, A., Snyder, G. E., Selvin, P. R. & Bezannilla, F. Atomic scale movement of the voltage-sensing region in a potassium channel measured via spectroscopy. *Nature* **402**, 809–813 (1999).
22. Hayden, F. G., Rollins, B. S. & Hay, A. J. Anti-influenza virus activity of the compound LY253963. *Antiviral Res.* **14**, 25–38 (1990).
23. Krug, R. M., Alonso-Caplen, F. V., Julkunen, I. & Katze, M. G. in *The Influenza Virus* 89–152 (Plenum, New York, 1989).
24. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770–774 (2000).
25. Ortmann, B. *et al.* A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* **277**, 1306–1309 (1997).
26. Stam, N. J., Spits, H. & Ploegh, H. L. Monoclonal antibodies raised against denatured HLA-B locus heavy chains permit biochemical characterization of certain HLA-C locus products. *J. Immunol.* **137**, 2299–2306 (1986).
27. Reits, E. A. J., Benham, A. M., Plougastel, B., Neeffes, J. & Trowsdale, J. Dynamics of proteasome distribution in living cells. *EMBO J.* **16**, 6087–6094 (1997).
28. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055–1069 (1976).
29. Jovin, T. M. & Vaz, W. L. Rotational and translational diffusion in membranes measured by fluorescence and phosphorescence methods. *Methods Enzymol.* **172**, 471–513 (1989).

Acknowledgements

We thank K. Döring, M. van Ham and K. Jalink for discussions, L. Oomen for help with the confocal microscope, R. Wubboldts for support with microinjection, P. Vlottes (ICN Holland) for a sample of ribavirin, and A. Benham and T. Schumacher for critically reading the manuscript.

Correspondence and requests for material should be addressed to J.N. (e-mail: jneefjes@nki.nl).

NAK is an I κ B kinase-activating kinase

Yuichiro Tojima[†], Atsushi Fujimoto[‡], Mireille Delhase[‡], Yi Chen[‡], Shigetugu Hatakeyama[§], Kei-ichi Nakayama[§], Yoko Kaneko[‡], Yuji Nimura[†], Noboru Motoyama^{*}, Kyoji Ikeda^{*}, Michael Karin[‡] & Makoto Nakanishi[¶]

^{*} Department of Geriatric Research, National Institute for Longevity Sciences, Obu, Aichi 474-8522, Japan
[†] Department of Surgery, Nagoya University Medical School, Shouwa-ku, Nagoya, Aichi 466-8550, Japan
[‡] Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, University of California San Diego, School of Medicine, La Jolla, California 92093-0636, USA
[§] Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyusyu University, Higashi-ku, Fukuoka 812-8582, Japan
^{||} CREST of Japan Science and Technology Corporation (JST), 3-1-1 Mardashi, Higashi-ku, Fukuoka 812-8582, Japan
[¶] Department of Biochemistry, Nagoya City University Medical School, Mizuho-ku, Nagoya, Aichi 467-8601, Japan

Phosphorylation of I κ B by the I κ B kinase (IKK) complex is a critical step leading to I κ B degradation and activation of transcription factor NF- κ B¹. The IKK complex contains two catalytic subunits, IKK α and IKK β , the latter being indispensable for NF- κ B activation by pro-inflammatory cytokines^{2–7}. Although IKK is activated by phosphorylation of the IKK β activation loop⁸, the physiological IKK kinases that mediate responses to extracellular stimuli remain obscure^{1,9}. Here we describe an IKK-related kinase, named NAK (NF- κ B-activating kinase), that can activate IKK through direct phosphorylation. NAK induces I κ B degradation and NF- κ B activity through IKK β . Endogenous NAK is activated by phorbol ester tumour promoters and growth factors, whereas

catalytically inactive NAK specifically inhibits activation of NF- κ B by protein kinase C- ϵ (PKC ϵ). Thus, NAK is an IKK kinase that may mediate IKK and NF- κ B activation in response to growth factors that stimulate PKC ϵ activity.

A portion of human NAK complementary DNA was amplified by polymerase chain reaction (PCR) using degenerate primers based on sequences common to IKK α and IKK β . A 240 base-pair (bp) PCR product was used to isolate a cDNA, whose sequence predicts a translation product of 730 amino acids with a calculated relative molecular mass of 84,000 (M_r 84K) (Fig. 1a). NAK exhibits 61% overall identity to IKK-i, which was recently described as a lipopolysaccharide

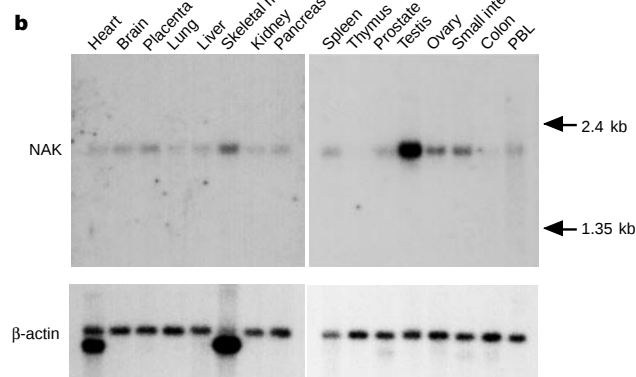
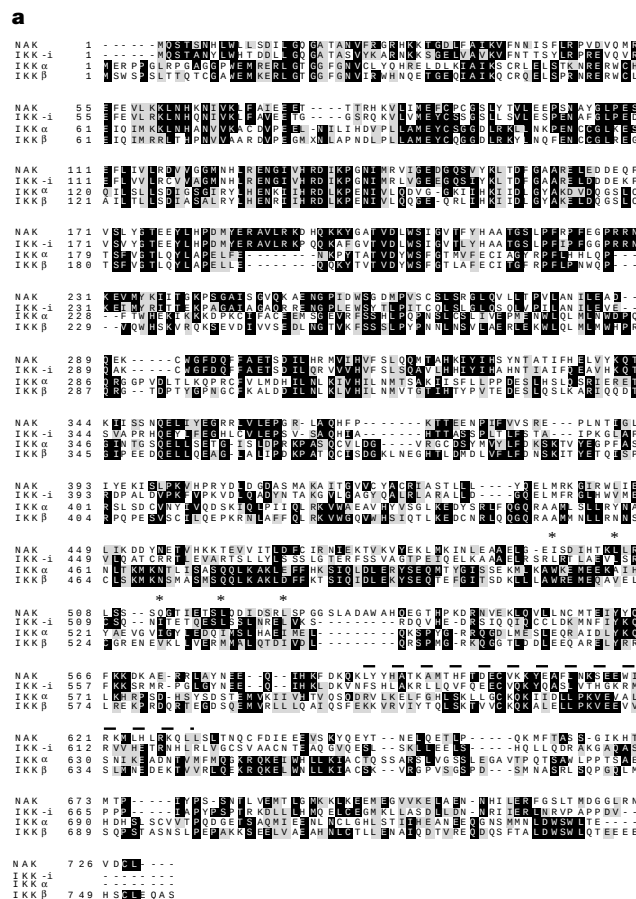


Figure 1 Structure and expression of NAK. **a**, Optimized alignment of the human NAK, IKK-i, IKK α and IKK β protein sequences is shown using single-letter codes. Identical and similar amino acids are black-boxed and shadow-boxed, respectively. The helix-loop-helix domain is marked by a dotted line and the leucine zipper is indicated by asterisk. **b**, Northern blot analyses of poly(A)⁺ RNA from the indicated human tissues with NAK (upper panel) and β -actin (lower panel) cDNA probes. PBL, peripheral blood leukocyte.

(LPS)-inducible I κ B kinase¹⁰, and its catalytic domain exhibits 30% identity to IKK α / β . Like IKK α / β , NAK contains leucine-zipper and helix-loop-helix motifs within its carboxy-terminal half. However, whereas IKK α and IKK β contain two serines in their activation loop, phosphorylation of which is required for activation^{1,8,11}, NAK contains glutamic acid (Glu 168) instead of one of these serines. A 2.2-kilobase (kb) NAK transcript is ubiquitously expressed with highest expression found in testis (Fig. 1b).

I κ B α phosphorylation at Ser 32 and Ser 36 is required to trigger its ubiquitination and subsequent degradation^{12–15}. Thus, we deter-

mined whether NAK induces I κ B α phosphorylation and ubiquitination *in vivo*. Expression of NAK alone induced phosphorylation of I κ B α at Ser 32 (Fig. 2a, lane 3), and co-expression of NAK and IKK β enhanced this modification (lane 6). Although overexpression of IKK β induces Ser 32 phosphorylation on its own¹ (data not shown), modest expression of IKK β alone was insufficient (lane 4). When FWD1(E3RS^{I κ B}/ β -TrCP), the specificity component of the SCF ubiquitin ligase for I κ B α ^{16,17}, was co-expressed with NAK, FWD1 bound phosphorylated I κ B α and stimulated its ubiquitination (lanes 7 and 10). Consistent with this, overexpression of NAK

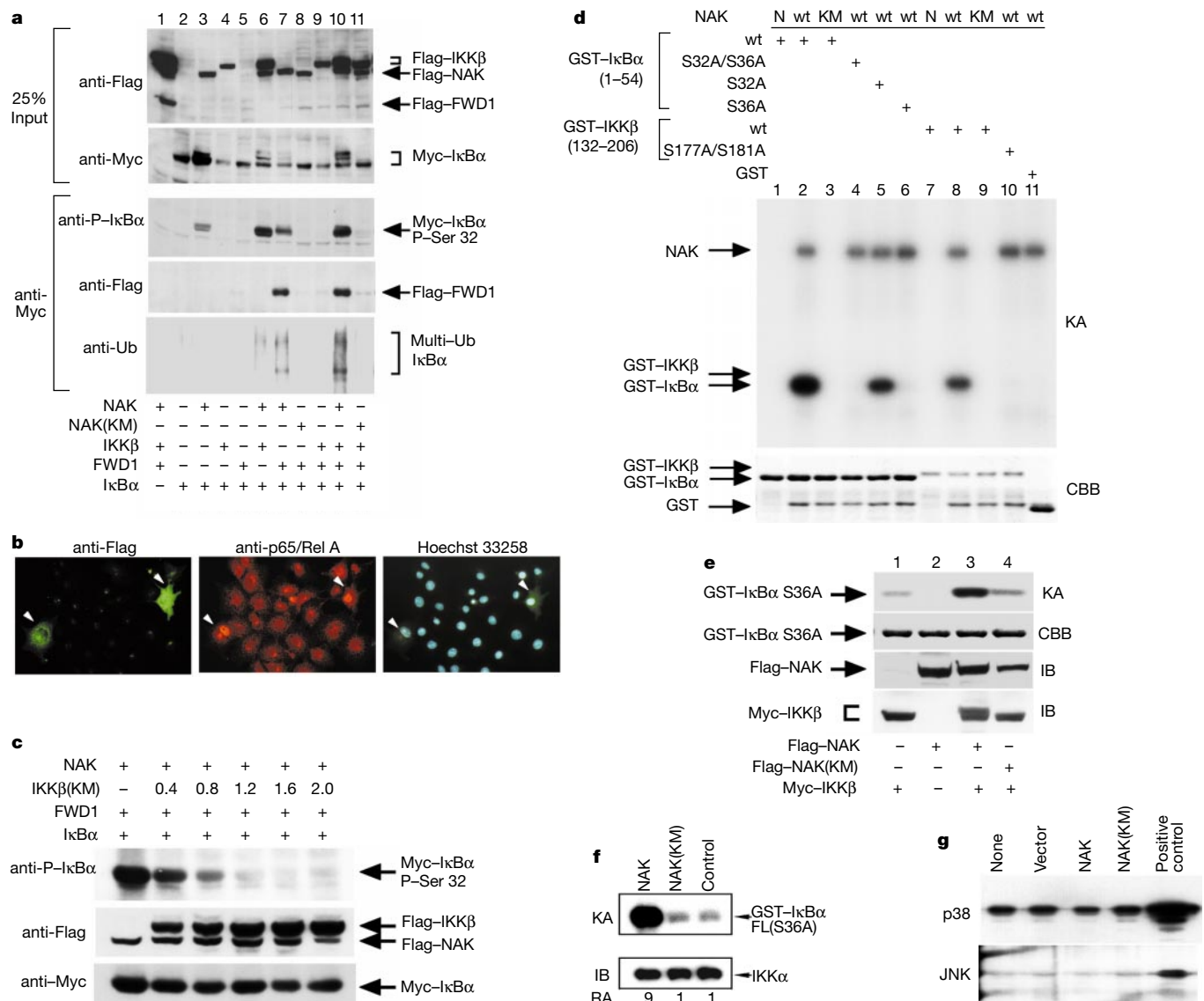


Figure 2 NAK induces phosphorylation and ubiquitination of I κ B α through activation of IKK β . **a**, 293T cells were transfected with the indicated expression vectors and lysed 36 h later. Anti-Myc immunoprecipitates were analysed by immunoblotting with antibodies specific to phospho-I κ B α (Ser 32) (anti-P-I κ B α ; NEB 9241S), ubiquitin (anti-Ub) or the Flag epitope. **b**, NAK facilitates nuclear translocation of NF- κ B. COS7 cells transfected with Flag-NAK were fixed and stained with anti-p65/RelA (middle), anti-Flag (left) and Hoechst33258 (right). Arrowheads, transfected cells. **c**, 293T cells were co-transfected with Flag-NAK, increasing amounts (μ g) of IKK β (KM), FWD1 and Myc-I κ B α . Phospho-I κ B α (Ser 32) was detected as described above (top). Expression levels of NAK, IKK β (middle) and I κ B α (bottom) were determined by immunoblotting. **d**, NAK phosphorylates I κ B α and IKK β *in vitro*. Myc-tagged wild-type NAK (wt) or its K38M mutant (KM) expressed in Sf9 cells were immunoprecipitated with anti-Myc or normal rabbit IgG (N), and assayed for phosphorylation of the indicated substrates. Top, autoradiography (KA);

bottom, Coomassie brilliant blue staining (CBB). **e**, NAK enhances IKK β kinase activity. After transfection with the indicated expression vectors, anti-Myc immunoprecipitates were assayed for phosphorylation (KA) of GST-I κ B α (1-54, S36A), whose presence is shown by CBB staining (CBB). Expression of NAK (third panel; IB) and IKK β (fourth panel; IB) was determined by immunoblotting. **f**, NAK directly activates IKK. IKK, chromatographically purified²⁹ from unstimulated HeLa cells, was incubated with purified NAK or NAK(KM) from Sf9 cells, or buffer (control), and was assayed for phosphorylation of GST-I κ B α (full-length, S36A; KA). The amount of IKK was determined by immunoblotting (IB) with anti-IKK α . I κ B α phosphorylation was quantified by phosphorimaging and used to calculate relative specific activities (RA). **g**, NAK does not stimulate JNK or p38MAPK activities. After transfection with NAK or NAK(KM), JNK and p38 activities were determined by immune-complex kinase assay as described⁹. Positive controls included osmotic shock (0.5 M NaCl) for p38 and ultraviolet irradiation (40 J m⁻²) for JNK.

resulted in nuclear translocation of the RelA(p65) NF- κ B subunit (Fig. 2b). However, NAK-induced phosphorylation of I κ B α was inhibited by co-expression of catalytically inactive IKK β (KM), in a dose-dependent manner (Fig. 2c, top), indicating that NAK may function upstream of IKK.

We also investigated whether NAK can phosphorylate I κ B α or IKK β directly. We expressed wild-type (wt) and catalytically inactive (KM) NAK in insect cells, as C-terminal Myc- and His-tagged proteins. Wild-type NAK, but not NAK(KM), phosphorylated GST-I κ B α (1–54) (Fig. 2d, compare lanes 2 and 3). We also observed autophosphorylation of NAK. GST-I κ B α with alanine at position 32 was phosphorylated by NAK (lane 5), but GST-I κ B α with alanine at position 36 was not (lane 6). Using similar mutants of I κ B β we found that NAK only phosphorylates Ser 23, and not Ser 19 (data not shown). Wild-type NAK also phosphorylated GST-IKK β (132–206) but NAK(KM) did not (compare lanes 8 and 9). However, wild-type NAK did not phosphorylate an alanine-substituted mutant GST-IKK β (S177A/S181A) (lane 10). Therefore, NAK can phosphorylate either or both serines in the activation loop of IKK β . Importantly, wild-type NAK stimulated the I κ B α kinase activity associated with IKK β , without changing IKK β expression level, but NAK(KM) did not (Fig. 2e). Wild-type NAK also reduced the electrophoretic mobility of IKK β (Fig. 2a, compare lanes 10 and 11, and Fig. 2e, compare lanes 3 and 4). As phosphatase treatment reversed this mobility shift (data not shown), it is probably due to IKK β phosphorylation.

As phosphorylation of I κ B α at both Ser 32 and Ser 36 is required for induction of I κ B α degradation^{15–17}, NAK is unlikely to function as an I κ B α kinase. Instead it is more likely to function as an IKK

kinase. To establish this concept, we examined whether recombinant NAK expressed in insect cells can activate the physiological IKK complex purified from unstimulated HeLa cells. Wild-type NAK, but not NAK(KM), enhanced IKK activity, as measured by phosphorylation of GST-I κ B α (full length; S36A) (Fig. 2f). Nck-interacting ste20 kinase (NIK) and mitogen-activated protein kinase kinase kinase-1 (MEKK-1) can activate c-Jun amino-terminal kinase and/or the *M*_r 28K subunit of mitogen-activated protein kinase (p38 MAPK) when overexpressed⁹. However, neither p38MAPK nor JNK was activated by overexpression of NAK (Fig. 2g). Thus, NAK may be a specific upstream regulator of IKK.

Next, we tested whether NAK stimulates the transcriptional activity of NF- κ B through IKK. When modest and comparable levels of NAK and IKK β were expressed, NAK caused approximately a sixfold increase in luciferase expression, whereas the effect of IKK β alone was weaker (Fig. 3a). Neither NAK(KM) nor IKK β (KM) stimulated NF- κ B activity. NF- κ B activation by NAK was further confirmed by electrophoretic mobility shift assay (EMSA) (Fig. 3b). Overexpression of IKK β , which results in its activation through autophosphorylation⁹, also enhanced NF- κ B DNA-binding activity (Fig. 3b) and caused a sevenfold increase in luciferase expression, which was not inhibited by NAK(KM) (Fig. 3c). In contrast, stimulation of NF- κ B-luciferase (Luc) expression by NAK was inhibited by IKK β (KM) (Fig. 3c), indicating that NAK acts upstream of IKK β . Consistent with this, activation of NF- κ B-Luc by NAK was reduced in *IKK α ^{-/-}* and *IKK β ^{-/-}* fibroblasts in comparison with fibroblasts derived from wild-type mouse embryo (Fig. 3d, left). The reduction in NF- κ B activation by NAK was more pronounced in *IKK β ^{-/-}* cells than in *IKK α ^{-/-}* cells, in agreement with

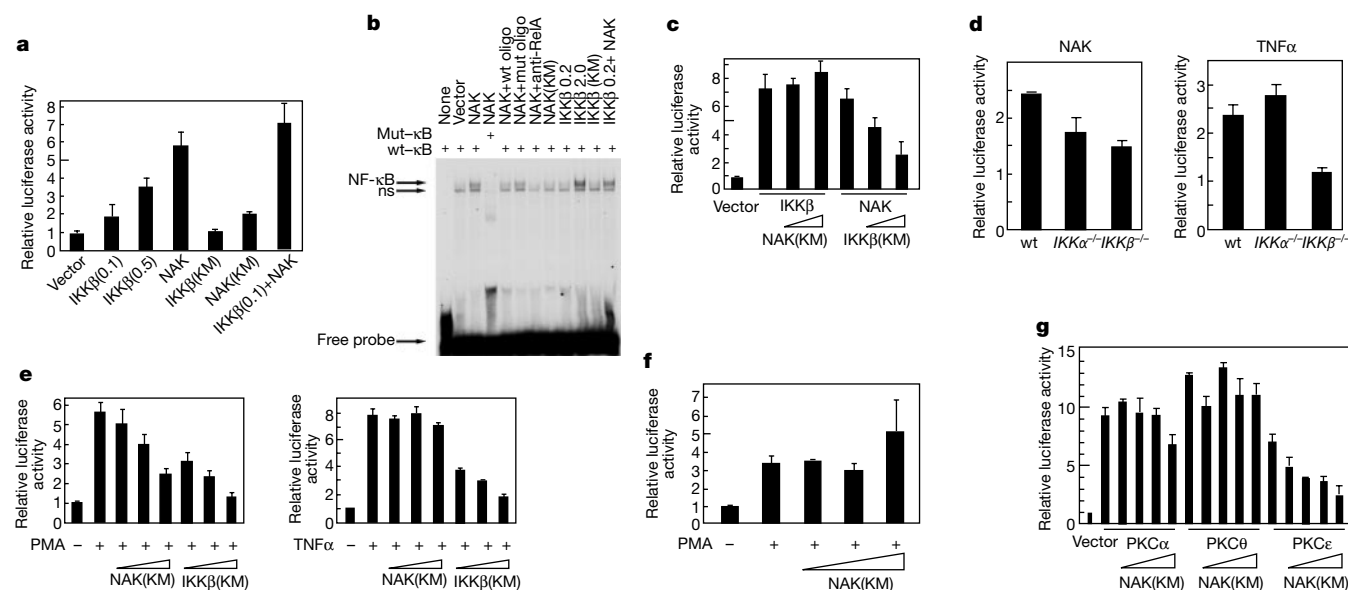


Figure 3 NAK activates NF- κ B. **a**, 293T cells were co-transfected with the indicated expression vectors, together with β -galactosidase expression vector driven by an actin promoter (pmivZ) and NF- κ B-Luc reporter. After 36 h, luciferase activities were determined, and normalized for β -galactosidase expression. The values are shown as mean $\pm$ s.d. ($n = 5$). IKK β (0.1) and IKK β (0.5) refer to the amounts of DNA (in μ g per plate) used for transfections. **b**, The indicated expression vectors were transfected into 293T cells and nuclear extracts were analysed by EMSAs with either wild-type (wt) or mutant (mut) κ B oligonucleotide probes. Some samples were pre-incubated with unlabelled wild-type or mutant κ B oligonucleotides, or anti-RelA antibody, as indicated. IKK β 0.2 and IKK β 2.0 refer to the amounts of DNA (in μ g per plate) used for transfections. ns, nonspecific. **c**, Increasing amounts of NAK(KM) (1.4 and 5.6 μ g per plate) or IKK β (KM) (0.5 and 2.0 μ g per plate) were co-transfected with wild-type IKK β or NAK, respectively, and the NF- κ B-Luc and actin- β -galactosidase reporters. Normalized luciferase activities were determined as described above. **d**, Fibroblasts derived from

wild-type, *IKK α ^{-/-}* or *IKK β ^{-/-}* mouse embryos were transfected with NAK and the 2 $\times$ NF- κ B-Luc reporters (NAK), or 2 $\times$ NF- κ B-Luc reporters followed by TNF α stimulation (TNF α). Luciferase activities were determined as described above. Results shown are averages of two experiments done in triplicate. **e**, The NF- κ B-Luc and actin- β -galactosidase reporters were co-transfected with increasing amounts of NAK(KM) (1.0, 2.0 and 3.0 μ g per plate) or IKK β (KM) (0.5, 1.0 and 2.0 μ g per plate). After 30 h, cells were treated with PMA (left panel) or TNF α (right panel) for 6 h. Luciferase activity was determined as described above. **f**, AP-1-Luc reporter was co-transfected with increasing amounts of NAK(KM) (1.0, 2.0 and 3.0 μ g per plate) and luciferase activities were determined after treatment with PMA. **g**, Expression vectors for several PKC isoforms (α , θ , ϵ)³⁰ were co-transfected with the NF- κ B-Luc and actin- β -galactosidase reporters and increasing amounts of NAK(KM) (0.5, 1.0, 2.0 and 3.0 μ g per plate). Luciferase activities were determined as described above.

the more important role of IKK β in NF- κ B activation by an external stimulus such as tumour necrosis factor- α (TNF α) (Fig. 3d, right)²⁻⁸.

Although NF- κ B is activated by a plethora of external stimuli¹⁸⁻²¹, it remains unclear how these stimuli activate a common downstream target, the IKK complex. To identify the pathway in which NAK operates, we tested the ability of NAK(KM) to interfere with NF- κ B activation. Induction of NF- κ B-Luc by a phorbol ester tumour promoter (PMA) was inhibited in a dose-dependent manner by either NAK(KM) or IKK β (KM) (Fig. 3e, left); therefore NAK may act downstream of protein kinase C (PKC)²². In contrast, induction of NF- κ B-Luc by TNF α , IL1- β , LPS and ionizing radiation were inhibited by IKK β (KM), but not by NAK(KM) (Fig. 3e, right, and data not shown). Nor did NAK(KM) inhibit induction of AP-1 transcriptional activity by PMA (Fig. 3f). Thus, the dominant inhibitory activity of NAK(KM) is stimulus- and NF- κ B-specific. Interestingly, NAK(KM) inhibited NF- κ B activation by PKC ϵ , but not by PKC α or PKC θ , indicating that it may be activated only by PKC ϵ or closely related isozymes (Fig. 3g). In this regard, it should be noted that PKC ϵ has been implicated in protection of cells from apoptosis^{23,24}. This anti-apoptotic function of PKC ϵ could be mediated through NAK, as NF- κ B is a well established anti-apoptotic factor²⁵.

We also examined the activity and regulation of endogenous NAK. Like the recombinant protein, endogenous NAK phosphorylates I κ B α only at Ser 36, and not at Ser 32 (Fig. 4a). Endogenous NAK was not present in IKK complexes purified from unstimulated or TNF α -stimulated HeLa cells (Fig. 4b). In addition NAK was not co-precipitated with IKK α , IKK β and IKK γ /NEMO in either PMA-treated or untreated cells (Fig. 4c). Therefore, NAK is not a component of the classical IKK complex¹ and does not become part of it upon PMA treatment. However, NAK activity was

stimulated by PMA with kinetics that matched those of IKK activation (Fig. 4d), and *de novo* protein synthesis was not required for this stimulation (data not shown). By comparison, there was only a minor and rather late increase in NAK activity in TNF α -stimulated cells, which did not correlate with IKK activation. While this work was being completed, NAK was described as a kinase (TBK1) capable of forming a ternary complex with TRAF2 and TANK when overexpressed²⁶. However, as kinase-defective NAK does not block physiological TRAF-dependent NF- κ B activation, it is questionable whether NAK is involved in TRAF-mediated signalling. We asked whether physiological stimuli, such as platelet-derived growth factor (PDGF) which activates PKC ϵ ²⁷, could activate NAK. PDGF induced a time-dependent increase in NAK activity, which was abrogated by Ro 31-8220, a specific PKC inhibitor (Fig. 4e). Hence, PDGF enhances NAK activity through PKC. In addition, NF- κ B activation by PDGF was inhibited by NAK(KM) (Fig. 4f). These results indicate that NAK is likely to function downstream of PKC ϵ or related isozymes and upstream of IKK in the signalling pathway through which PDGF or other growth factors stimulate NF- κ B activity.

In conclusion, our data suggest that although NAK has similar general organization to IKK α/β with especially high similarity within its kinase domain, it functions as an IKK kinase rather than an I κ B kinase. Although NAK can phosphorylate one of the two regulatory serines required for induction of I κ B degradation, this is insufficient for NF- κ B activation and NAK does not stably associate with other I κ B kinases. Although NAK may activate only a small fraction of IKK because of its relatively low abundance, it is possible that through a positive feedback mechanism NAK-activated IKK β can phosphorylate an adjacent IKK subunit through intramolecular *trans*-autophosphorylation and thus propagate a higher level of IKK activation¹. □

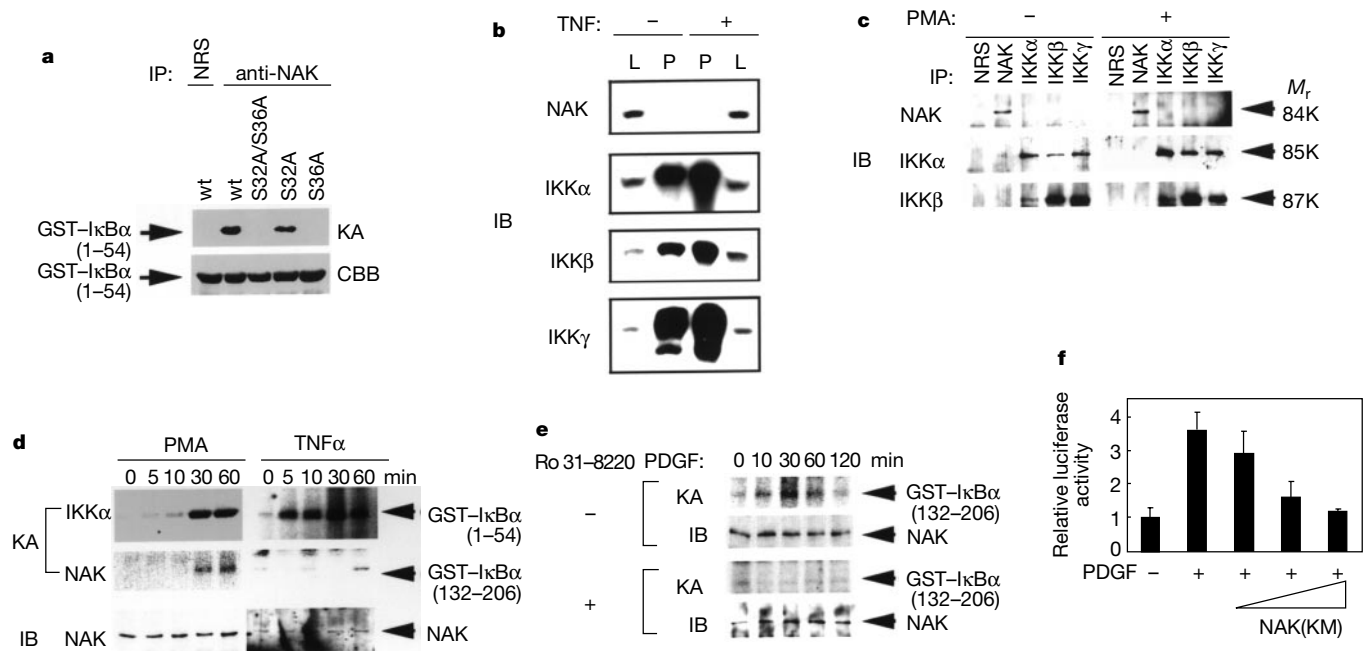


Figure 4 Endogenous NAK is activated by PMA and PDGF and is not part of the IKK complex. **a**, Serum-starved HeLa cells were treated with PMA for 30 min. Cell lysates (2 mg) were subjected to immune-complex kinase assay with the indicated I κ B α or GST-IKK β as substrates after immunoprecipitation with anti-NAK. NRS, normal rabbit serum. **b**, S-100 cytoplasmic extract (L) and purified IKK complex (P) from unstimulated (-) or TNF α -stimulated (+) HeLa cells were immunoblotted with antibodies specific to NAK, IKK α , IKK β and IKK γ . **c**, Unstimulated and PMA-stimulated HeLa cell extracts were subjected to immunoprecipitation-immunoblot analysis using the indicated antibodies. **d**, Serum-

starved HeLa cells were treated with PMA or TNF α and lysed at the indicated times. IKK and NAK activities were determined by immune-complex kinase assays using GST-I κ B α or GST-IKK β as substrates. **e**, Serum-starved A172 cells were treated with 100 ng ml⁻¹ PDGF in the presence or absence of Ro 31-8220 (5 μ M), and lysed at the indicated times. NAK activity was determined by immune-complex kinase assay. **f**, Increasing amounts of NAK(KM) (2.0, 3.0 and 4.0 μ g per plate) were transfected with NF- κ B-Luc and actin- β -galactoside reporters into A172 cells and, 30 h after transfection, cells were treated with PDGF for additional 6 h. Luciferase activity was determined as described above.

Methods

Molecular cloning of NAK

Degenerate primers, 5'-AT(TCA)AT(TCA)CA(TC)(CA)GNGA(TC)ATCAAACC-3' and 5'-GG(AGT)ATNCCNA(AG)(TC)TT(TC)TCAAAGAT-3', based on conserved motifs in the kinase domains of IKK α and IKK β , were used to synthesize cDNAs by PCR with reverse transcription (RT-PCR) of RNA derived from MDAH041 cells. Additional 5' sequences were obtained by 5' rapid amplification of cDNA ends (RACE), and 3' sequences were obtained by screening a human fetal brain cDNA library (Stratagene).

Preparation of recombinant NAK protein and anti-NAK antibody

Baculoviruses expressing Myc- and His $\times$ 6-tagged NAK (Myc/His-NAK) were generated by co-transfection of pVL1393-NAK with linearized baculovirus DNA (BaculoGold; Pharmingen) into Sf9 cells. After amplification of the virus, cell lysates were prepared 48 h after infection and Myc/His-NAK was purified on ProBond Resin (Invitrogen). GST-NAK protein expressed in *Escherichia coli* was purified on glutathione-Sepharose beads (Pharmacia), and used to immunize rabbits. Antisera were precleared on HiTrap NHS column coupled with GST protein and then affinity purified on HiTrap NHS column coupled with GST-NAK protein. The protein concentration was adjusted to 0.6 mg ml⁻¹.

Kinase assay

Wild-type NAK and NAK(KM) proteins expressed in Sf9 cells were immunoprecipitated with anti-Myc antibody (MBL). The kinase activity was determined at 30 °C for 30 min in a 30 μ l reaction mixture containing 50 mM HEPES (pH 8.0), 10 mM MgCl₂, 2.5 mM EGTA, 1 mM DTT, 10 μ M β -glycerophosphate, 1 mM NaF, 0.1 mM Na₃VO₄, 0.1 mM PMSF, 10 μ M ATP and 185 kBq [γ -³²P]ATP (222 TBq mmol⁻¹; NEM). Reaction products were separated by SDS-PAGE, and phosphorylated proteins were detected by autoradiography. For immune-complex kinase assay, serum-starved HeLa and A172 cells (2 $\times$ 10⁶ cells) were treated with PMA (20 nM) and PDGF (50 ng ml⁻¹ PDGF-AA plus 50 ng ml⁻¹ PDGF-BB), respectively. The treated cells were then lysed in an IP-kinase buffer (50 mM HEPES (pH 8.0), 150 mM NaCl, 25 mM EGTA, 1 mM EDTA, 0.1% Tween 20 and 10% glycerol) containing a cocktail of protease inhibitors (20 μ g ml⁻¹ soybean trypsin inhibitor, 2 μ g ml⁻¹ aprotinin, 5 μ g ml⁻¹ leupeptin and 100 μ g ml⁻¹ PMSF) and phosphatase inhibitors (50 mM NaF and 0.1 mM Na₃VO₄). The cell lysates (2 mg protein) were immunoprecipitated with anti-NAK antibody or anti-IKK α antibody (Santa Cruz; M280), and the immunoprecipitates were assayed for kinase activity using GST-IkBa or GST-IKK β as substrates. The immunoprecipitates were also subjected to western blotting to determine the amount of precipitated proteins.

Transfections, immunoblot analysis, reporter assays, EMSAs and immunostaining

293T cells on 35-mm plates were transfected with pcDNA3-Flag-NAK, pCR3-Flag-IKK β , pcDNA3-Flag-FWD1 or pcDNA3-myc-IkBa using FuGENE 6 transfection kit (Boehringer). After 36 h, cells were lysed in IP-kinase buffer. Immunoprecipitates were separated by SDS-PAGE and analysed by western blotting.

Luciferase activity was determined 36 h after transfection, using an enhanced luciferase assay kit and a luminometer. All transfections included pmiwZ²⁸ to normalize transfection efficiencies based on β -galactosidase activity. The amounts of NAK and IKK β expressed in 293T cells were determined by western blotting using anti-Flag antibody (M5, IBI).

EMSAs were performed on lysates of transiently transfected 293T cells. Binding of DNA by NF- κ B was assayed using a ³²P-labelled κ B probe.

Immunostaining of COS-7 cells transfected with expression vectors of Flag-NAK was performed as described previously¹⁷.

Received 19 November 1999; accepted 14 February 2000.

- Karin, M. The beginning of the end: IkB kinase (IKK) and NF- κ B activation. *J. Biol. Chem.* **274**, 27339–27342 (1999).
- Hu, Y. *et al.* Abnormal morphogenesis but intact IKK activation in mice lacking the IKK α subunit of IkB kinase. *Science* **284**, 316–320 (1999).
- Takeda, K. *et al.* Limb and skin abnormalities in mice lacking IKK α . *Science* **284**, 313–316 (1999).
- Tanaka, M. *et al.* Embryonic lethality, liver degeneration, and impaired NF- κ B activation in IKK- β -deficient mice. *Immunity* **10**, 421–429 (1999).
- Li, Q., Van Antwerp, D., Mercurio, F., Lee, K.-F. & Verma, I. M. Severe liver degeneration in mice lacking the IkB kinase 2 gene. *Science* **284**, 321–325 (1999).
- Li, Q. *et al.* IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**, 1322–1328 (1999).
- Li, Z. W. *et al.* The IKKbeta subunit of IkappaB kinase (IKK) is essential for nuclear factor kappaB activation and prevention of apoptosis. *J. Exp. Med.* **189**, 1839–1845 (1999).
- Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. Positive and negative regulation of IkappaB kinase activity through IKKbeta subunit phosphorylation. *Science* **284**, 309–313 (1999).
- Karin, M. & Delhase, M. JNK or IKK, AP-1 or NF- κ B, which are the targets for MEK kinase 1 action. *Proc. Natl Acad. Sci. USA* **95**, 9067–9069 (1998).
- Shimada, T. *et al.* IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases. *Int. Immunol.* **8**, 1357–1362 (1999).
- Ling, L., Cao, Z. & Goeddel, D. V. NF- κ B-inducing kinase activates IKK- α by phosphorylation of Ser-176. *Proc. Natl Acad. Sci. (USA)* **95**, 3792–3797 (1998).
- Brockman, J. A. *et al.* Coupling of a signal response domain in IkBa to multiple pathways for NF- κ B activation. *Mol. Cell. Biol.* **15**, 2809–2818 (1995).
- Brown, K., Gerstberger, S., Carlson, L., Franzoso, G. & Siebenlist, U. Control of IkB- α proteolysis by site-specific, signal-induced phosphorylation. *Science* **267**, 1485–1488 (1995).
- Traencker, E. B.-M. *et al.* Phosphorylation of human IkB- α on serines 32 and 36 controls IkB- α proteolysis and NF- κ B activation in response to diverse stimuli. *EMBO J.* **14**, 2876–2883 (1995).

- DiDonato, J. A. *et al.* Mapping of the inducible IkB phosphorylation sites that signal its ubiquitination and degradation. *Mol. Cell. Biol.* **16**, 1295–1304 (1996).
- Yaron, A. *et al.* Identification of the receptor component of the IkBa-ubiquitin ligase. *Nature* **396**, 590–594 (1998).
- Hatakeyama, S. *et al.* Ubiquitin-dependent degradation of IkBa is mediated by a ubiquitin ligase Skp1/Cullin/F-box protein FWD1. *Proc. Natl Acad. Sci. USA* **96**, 3859–3863 (1999).
- Baeuerle, P. A. & Baltimore, D. NF- κ B: ten years after. *Cell* **87**, 13–20 (1996).
- Verma, I. M., Stevenson, J. K., Schwarz, E. M., Van Antwerp, D. & Miyamoto, S. Rel/NF- κ B family: intimate tales of association and dissociation. *Genes Dev.* **9**, 2723–2735 (1995).
- Baldwin, A. S. The NF- κ B and IkB proteins: new discoveries and insights. *Annu. Rev. Immunol.* **14**, 649–683 (1996).
- Barnes, P. J. & Karin, M. Nuclear factor- κ B: a pivotal transcription factor in chronic inflammatory diseases. *N. Engl. J. Med.* **336**, 1066–1071 (1997).
- Nishizuka, Y. Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. *Science* **258**, 607–614 (1992).
- Gray, M. O., Karlner, J. S. & Mochly-Rosen, D. A selective e-protein kinase C antagonist inhibits protection of cardiac myocytes from hypoxia-induced cell death. *J. Biol. Chem.* **272**, 30945–30951 (1997).
- Mayne, G. C. & Murray, A. W. Evidence that protein kinase C ϵ mediates phorbol ester inhibition of calphostin C- and tumor necrosis factor- α -induced apoptosis in U937 histiocytic lymphoma cells. *J. Biol. Chem.* **273**, 24115–24121 (1998).
- Liu, Z. G., Hsu, H., Goeddel, D. V. & Karin, M. Dissection of TNF receptor 1 effector functions: JNK activation is not linked to apoptosis while NF- κ B activation prevents cell death. *Cell* **87**, 565–576 (1996).
- Pomerantz, J. L. & Baltimore, D. NF- κ B activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **18**, 6694–6704 (1999).
- Moriya, S. *et al.* Platelet-derived growth factor activates protein kinase C ϵ through redundant and independent signaling pathways involving phospholipase C γ or phosphatidylinositol 3-kinase. *Proc. Natl Acad. Sci. USA* **93**, 151–155 (1996).
- Suemori, H. *et al.* A mouse embryonic stem cell line showing pluripotency of differentiation in early embryos and ubiquitous beta-galactosidase expression. *Cell Differ. Dev.* **29**, 181–186 (1990).
- DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive IkB kinase that activates the transcription factor NF- κ B. *Nature* **388**, 548–554 (1997).
- Werlen, G., Jacinto, E., Xia, Y. & Karin, M. Calcineurin preferentially synergizes with PKC-theta to activate JNK and IL-2 promoter in T lymphocytes. *EMBO J.* **17**, 3101–3111 (1997).

Acknowledgements

We thank D. Rothwarf for discussions and help with the IKK activation and association experiments, K. Tsujimura, Y. Obata and T. Takahashi for generating anti-NAK antibody, and M. Shirane for immunohistochemical study. This work was supported by the Ministry of Education, Science, Sports, and Culture of Japan, and the Ministry of Health and Welfare of Japan. Research in M.K. lab was supported by grants from the National Institute of Health (NIH) and the Department of Energy. M.K. is an American Cancer Society Research Professor.

Correspondence and requests for materials should be addressed to M.N. (e-mail: mkt-naka@med.nagoya-cu.ac.jp). The NAK cDNA sequence has been deposited with GenBank (accession number AF174536).

AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27^{kip1}

René H. Medema^{*†}, Geert J. P. L. Kops^{†‡}, Johannes L. Bos[‡] & Boudewijn M. T. Burgering[‡]

** Jordan Laboratory G03-647, Department of Hematology and ‡ Department of Physiological Chemistry and Centre for Biomedical Genetics, University Medical Center, Utrecht 3584 CG, The Netherlands*

† These authors contributed equally to this work

The Forkhead transcription factors AFX, FKHR and FKHR-L1 are orthologues of DAF-16, a Forkhead factor that regulates longevity in *Caenorhabditis elegans*^{1–3}. Here we show that overexpression of these Forkhead transcription factors causes growth suppression in a variety of cell lines, including a Ras-transformed cell line and a cell line lacking the tumour suppressor PTEN. Expression of AFX blocks cell-cycle progression at phase G₁, independent of

functional retinoblastoma protein (pRb) but dependent on the cell-cycle inhibitor p27^{kip1}. Indeed, AFX transcriptionally activates p27^{kip1}, resulting in increased protein levels. We conclude that AFX-like proteins are involved in cell-cycle regulation and that inactivation of these proteins is an important step in oncogenic transformation.

In *C. elegans*, DAF-16 is inhibited by a pathway consisting of proteins that are orthologues of the insulin receptor, phosphatidylinositol-3-OH kinase (PI(3)K) and protein kinase B (PKB)/c-Akt²⁻⁴. The same pathway exists in vertebrates, where AFX, FKHR and FKHR-L1 are directly phosphorylated and inhibited by PKB⁵⁻⁹. Additionally, AFX-mediated transcription is negatively regulated through Ras/Ral signalling⁵. As PI(3)K/PKB and Ras signalling are both important in cell-cycle progression through G₁ phase¹⁰⁻¹³, we hypothesized that the members of the AFX family of Forkhead transcription factors might regulate aspects of G₁ progression. Therefore, we examined the effect of ectopic expression of these AFX-like proteins on cell proliferation in mammalian cells.

Transfection of A14 cells⁵ with AFX, FKHR or FKHR-L1 resulted in efficient growth suppression (Fig. 1a). When mutants of these proteins which lack PKB phosphorylation sites (AFX.A3, FKHR.A3 or FKHR-L1.A3) were introduced, they inhibited colony formation even more efficiently, to levels observed with the cyclin-dependent kinase (cdk) inhibitor p16^{ink4a} (Fig. 1a). However, a mutant of AFX lacking the DNA-binding domain (AFX.ΔDB) had no effect (Fig. 1a). Ectopic AFX and AFX.A3 were expressed at similar levels 48 h after transfection, but very little ectopic AFX or AFX.A3 could be detected in colonies that grew after 2 weeks of selection (Fig. 1a). AFX is regulated by PKB and Ras signalling, so

we examined the consequences of expression of AFX or AFX.A3 in the PTEN-deficient glioblastoma cell line U87MG, which contains high levels of PKB activity¹⁴, and in the RasV12-transformed Rat-1 cell line, RR2 (ref. 15). Expression of wild-type AFX caused a marked inhibition of colony formation in these transformed cell lines (Fig. 1b). Growth inhibition was even more pronounced when using AFX.A3 (Fig. 1b). We obtained similar data with another RasV12-transformed Rat-1 cell line, RR7 (ref. 15) (not shown).

Introduction of AFX into A14 cells resulted in accumulation of cells in the G₀/G₁ phase of the cell cycle (Fig. 1c), indicating that these cells are arrested in G₁. This change in cell-cycle distribution was even more clear when cells were treated with nocodazole after transfection so that cycling cells would accumulate in the G₂/M phase, unless they had been arrested in G₁ (Fig. 1c). Expression of AFX.ΔDB had no effect on the cell-cycle distribution (Fig. 1d), indicating that transcriptional activity is required for G₁ arrest. Insulin treatment overcame the G₁ arrest induced by AFX, but had very little effect on mock-transfected cells in the presence of 10% fetal calf serum (FCS) (Fig. 1d). However, insulin did not overcome the AFX.A3-induced cell-cycle arrest (Fig. 1d). This is consistent with the finding that insulin efficiently inhibits AFX-mediated transcription⁵, but transcriptional activation by AFX-A3 is only marginally affected (data not shown). Fluorescence-activated cell sorting (FACS) analysis of other cell lines, in particular tumour-derived cell lines, demonstrated that expression of AFX, AFX.A3, FKHR-L1 or FKHR-L1.A3 in U87MG, RR2, RR7 or U2OS osteosarcoma cells caused G₁ arrest (Fig. 1c and data not shown). Observation of the percentage of cells with less than 2N DNA content indicated only minor increases in the number of apoptotic

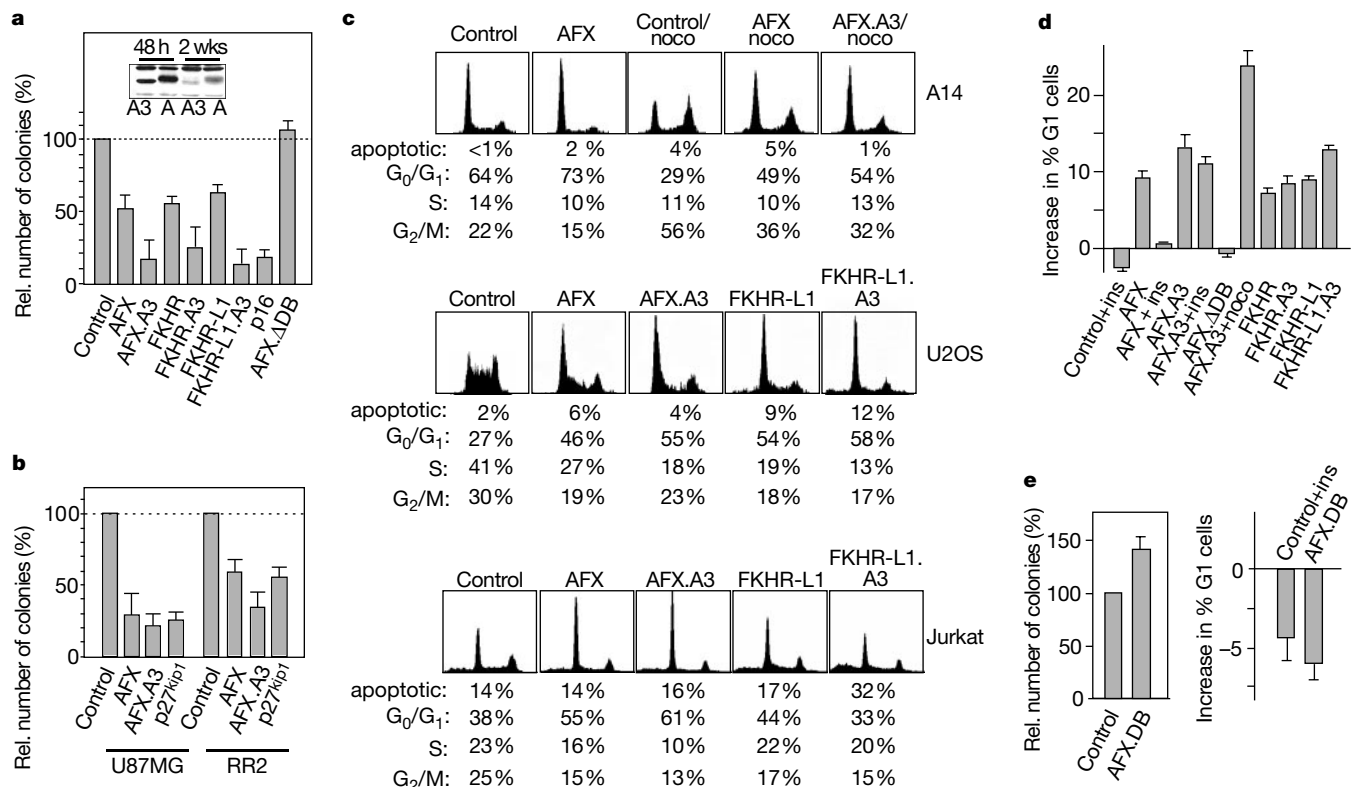


Figure 1 Expression of AFX causes arrest at the G₁ phase of the cell cycle. Colony formation assays in A14 cells (**a**), U87MG cells and RR2 (**b**). Cells were transfected with an expression plasmid for wild-type or mutant AFX, FKHR or FKHR-L1 in combination with pBabe.puro. Puromycin-resistant colonies were scored after two weeks of selection. Levels of ectopically expressed AFX (A) and AFX.A3 (A3) in total lysates at 48 h and two weeks after transfection were detected by western blot (inset). **c**, Cell-cycle profiles of

transfected A14, U2OS or Jurkat cells expressing the indicated AFX/FKHR.L1 proteins or respective mutants. **d**, Absolute increase in percentage of A14 cells in the G₀/G₁ phase upon expression of the indicated proteins. **e**, Increase in colony formation and a decrease in G₁ content in cells expressing the AFX.DB mutant. The data in **d** and **e** are the averages of three independent experiments. Where indicated cells were treated with nocodazole (noco) or insulin (ins) during the last 16 h before collection.

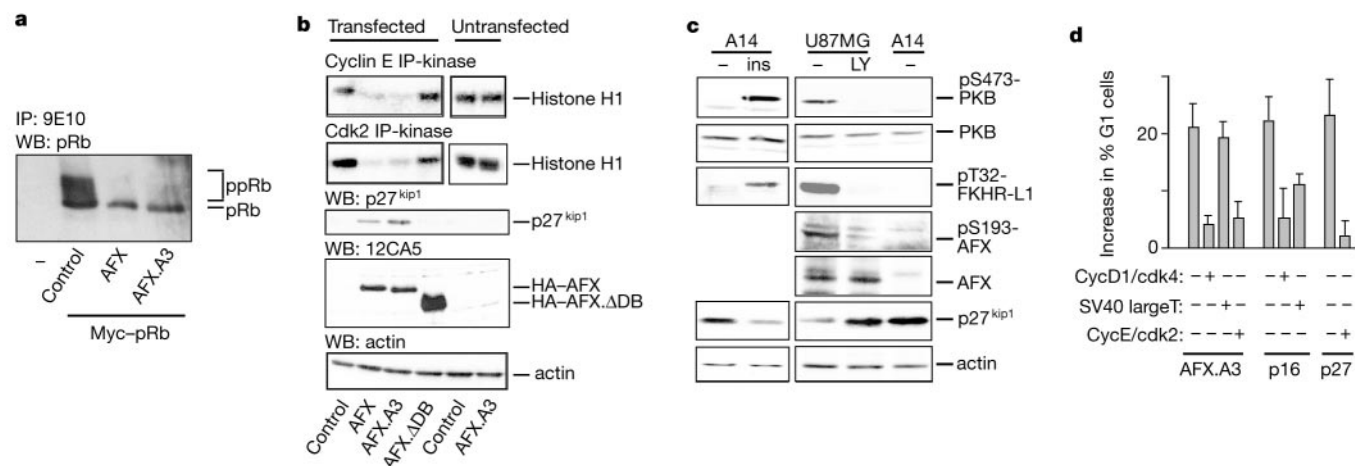


Figure 2 AFX expression induces up-regulation of p27^{kip1}. **a**, Myc-tagged pRb was transfected in combination with the indicated AFX mutants. Exogenous pRb was immunoprecipitated using anti-myc monoclonal antibodies (9E10) and the phosphorylation state was analysed on western blots (WB) using anti-pRb (Pharmingen). The first lane represents untransfected A14 cells. ppRb, hyperphosphorylated pRb. **b**, Transfected and untransfected A14 cells were separated by MACS, lysed and used for analysis of expression of p27^{kip1}, HA-tagged AFX, AFX.A3 and AFX.ΔDB. Actin expression is the loading control. Cyclin-E- and cdk2-associated kinase activity was analysed by *in vitro*

kinase reaction²². **c**, A14 cells were treated with insulin (ins) and U87MG cells were treated with LY294002 (LY). Treatment was performed for 16 h, after which cells were lysed and total lysates were probed for expression of PKB, phosphorylated PKBpS473, phosphorylated AFXpS193, phosphorylated FKHR-L1pT32, and expression of p27^{kip1} and actin as additional loading control. **d**, A14 cells were transfected with the indicated expression plasmids, treated with nocodazole for the last 16 h after transfection and analysed by flow cytometry. The data are the averages of three independent experiments.

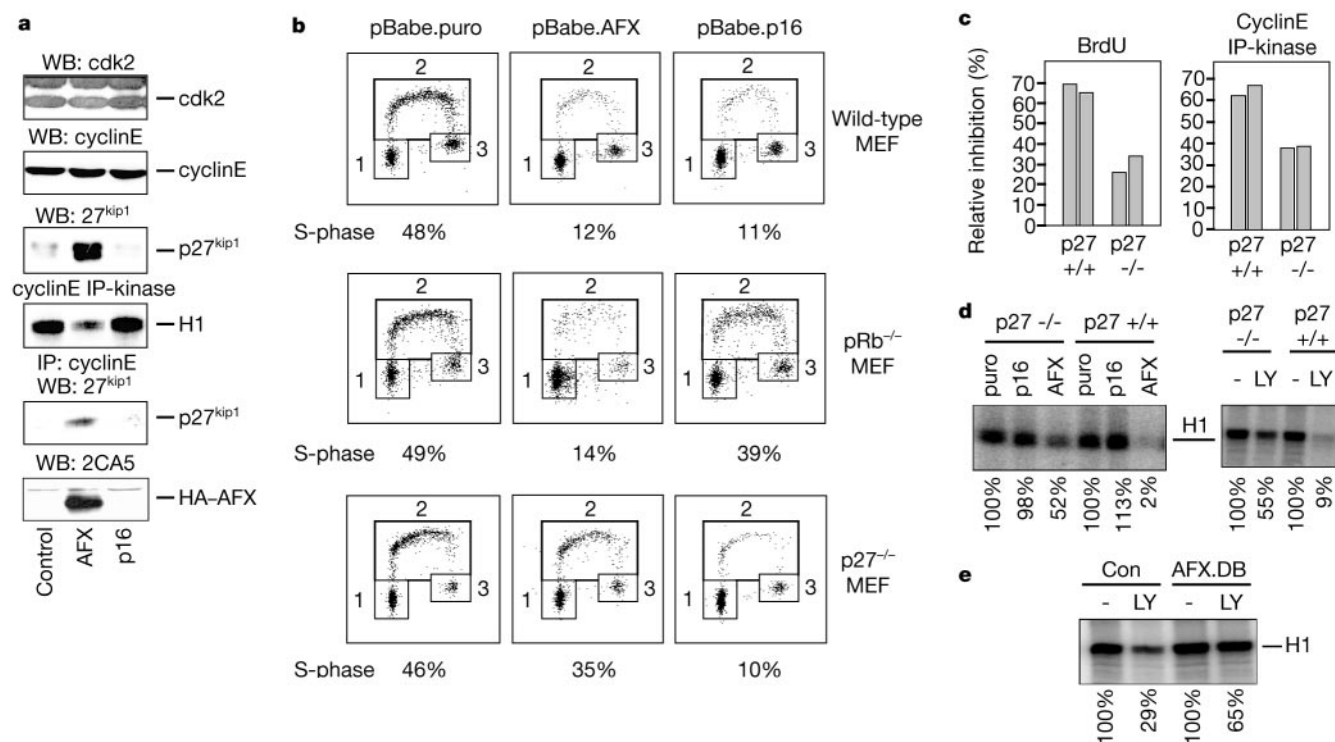


Figure 3 AFX-induced cell-cycle arrest occurs independent of pRb but is dependent on p27^{kip1}. **a**, Expression of exogenous AFX, induction of p27^{kip1}, expression of cyclin E and cdk2, inhibition of cyclin E-associated kinase activity and association of p27^{kip1} to cyclin E, in immortalized wild-type MEFs upon infection with an AFX- or p16^{ink4a}-encoding retrovirus. **b**, Dot-plots of BrdU-fluorescence versus DNA content of immortalized wild-type, pRb-deficient or p27^{kip1}-deficient MEFs infected with a control virus (pBabe.puro), a retrovirus encoding AFX (pBabe.AFX) or p16^{ink4a} (pBabe.p16) at 48 h after infection. Areas 1, 2 and 3 represent G₀/G₁, S and G₂/M, respectively. **c**, Primary cultures of wild-type and p27-deficient MEFs were infected with the AFX-encoding retrovirus. Inhibition of BrdU incorporation (left) and cyclin-E-associated kinase activity (right) relative to cultures

infected with the control retrovirus (pBabe.puro, puro) were determined at 48 h after infection. Each bar is an independent experiment. **d**, Inhibition of cyclin-E-associated kinase activity by ectopic AFX or LY294002 in immortalized wild-type or p27^{-/-} MEFs. Cells were infected with a control, p16^{ink4a}- or AFX-encoding retrovirus (left), or treated with LY294002 for 24 h (right), and cyclin-E-associated kinase activity was determined by *in vitro* kinase assays. Percentages below each lane indicate the relative activity compared to the control samples. **e**, A14 cells were transfected with a control plasmid or a plasmid encoding AFX-DB and treated overnight with LY294002 or left untreated. Transfected cells were isolated by MACS and cyclin E-associated kinase activity was determined by *in vitro* kinase assays²². H1, histone H1.

cells. For comparison, we studied Jurkat T cells, which become apoptotic upon introduction of the FKHR-L1.A3 mutant⁶. However, increased apoptosis was only seen with this particular mutant; all other AFX/FKHR.L1 expression plasmids tested caused a G₁ arrest (Fig. 1c). Thus, although apoptosis may be affected, the G₁ arrest seems to be a more general response to all AFX-like proteins.

To investigate whether inhibition of endogenous AFX-like activity would affect the proliferative capacity of A14 cells, we expressed a mutant of AFX encoding only the DNA-binding domain (AFX.DB). This mutant is unable to transactivate (data not shown), but is likely to compete with endogenous AFX-like proteins for binding to DNA. Interestingly, expression of AFX.DB resulted in a significant increase in colony formation (greater than 40%) and a decrease in the percentage of cells in G₁, similar to that observed upon addition of insulin (Fig. 1e). These data show that interference with endogenous AFX-like activity results in a proliferative advantage in these cells.

To investigate the mechanism underlying the AFX-induced G₁ arrest in more detail, we transfected A14 cells with myc-tagged pRb in combination with AFX or AFX.A3. Expression of AFX or AFX.A3 caused an inhibition of pRb-phosphorylation, indicating that AFX can arrest cell-cycle progression before the G₁ restriction point (Fig. 2a). To analyse the expression and activity of endogenous cell-cycle regulators, we separated the transfected and untransfected A14 cells by magnetic-activated cell sorting (MACS). Exogenous AFX and AFX.A3 were detected in the transfected cells, but no protein was seen in the untransfected cells, indicating that separation was effective (Fig. 2b). We found that the kinase activity of cyclin E and cdk2 immunoprecipitates were significantly inhibited in cells expressing AFX or AFX.A3 (Fig. 2b), although cyclin E and cdk2 protein levels were not decreased (data not shown). Expression of the cdk inhibitor p21^{waf1} was unchanged upon introduction of AFX

proteins (data not shown). In contrast, expression of AFX or AFX.A3 caused an increase in the protein levels of the cdk inhibitor p27^{kip1} (Fig. 2b), suggesting that AFX-induced upregulation of p27^{kip1} is responsible for the observed inhibition of cyclin E/cdk2 complexes. These effects were not seen with the AFX.ΔDB mutant (Fig. 2b).

Next we examined the effect of PKB activation or inactivation on the expression of p27^{kip1} and phosphorylation of endogenous AFX and FKHR-L1. Insulin stimulation of A14 cells resulted in activation of PKB and a reduction in p27^{kip1} expression (Fig. 2c). We were unable to detect endogenous AFX in A14 cells, but saw induction of FKHR-L1 phosphorylation in response to insulin (Fig. 2c). Conversely, inhibition of PKB in U87MG cells by addition of the PI(3)K inhibitor LY294002 caused a reduction in AFX and FKHR-L1 phosphorylation and a concomitant increase in p27^{kip1} expression (Fig. 2c). Therefore, PKB-mediated phosphorylation of AFX-like Forkhead factors is coupled with a decrease in p27^{kip1} levels in both of these cell lines.

We reasoned that if induction of p27^{kip1} levels is responsible for the AFX-induced G₁ arrest, co-expression of cyclin/cdk complexes should titrate out the induced p27^{kip1} and rescue cells from the cell-cycle arrest. Indeed, combinations of cyclin D/cdk4 or cyclin E/cdk2 could fully override the AFX.A3-induced cell-cycle arrest (Fig. 2d). In addition, we examined the effect of co-expression of the SV40 large T oncoprotein, which binds and inactivates pRb and p53. Co-expression of SV40 large T did not rescue the AFX.A3-induced G₁ arrest, although it could override a p16^{ink4a}-induced cell cycle arrest (Fig. 2d).

Cells devoid of functional pRb no longer require cyclin-D-associated kinase activity, but do need cyclin-E-associated kinase activity to initiate S phase¹⁶. Thus, if p27^{kip1} mediates the AFX-induced cell-cycle block, pRb-deficient cells should arrest in

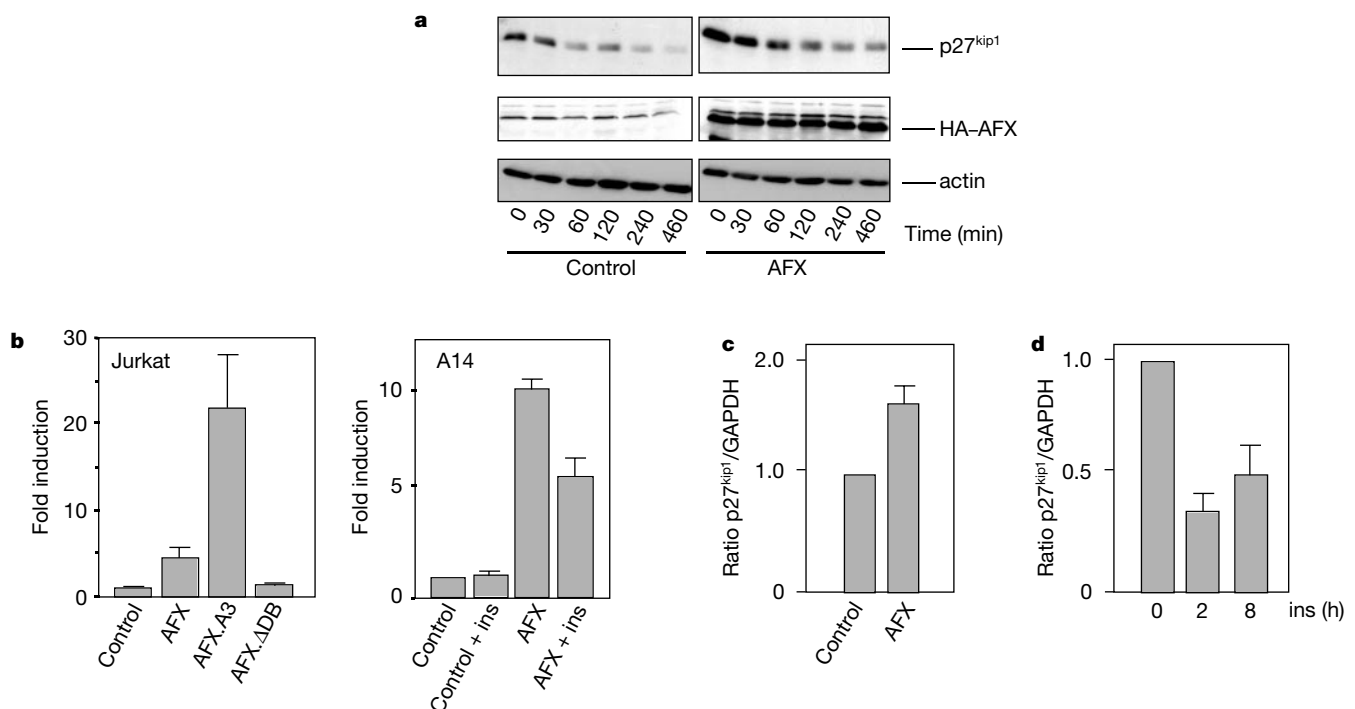


Figure 4 AFX regulates p27^{kip1} transcription. **a**, Wild-type MEFs were infected with either the control or the AFX-expressing (AFX) retrovirus. Forty-eight hours post-infection, cells were treated with cycloheximide and p27^{kip1} levels were followed by western blotting of total cellular lysates. **b**, Jurkat cells (left) and A14 cells (right) were transiently transfected with the p27GL-1609 construct with or without the various AFX plasmids and luciferase activity was measured. A14 cells were left untreated (–) or treated with insulin (+) for 16 h.

Luciferase levels were corrected for LacZ expression. **c**, Total RNA (20 μg) was isolated from AFX- or control-infected MEFs, blotted onto nitrocellulose and probed for the presence of p27^{kip1} messenger. Levels were corrected for GAPDH. **d**, Total RNA (20 μg) from A14 cells treated with insulin for the indicated periods was probed for p27^{kip1} messenger levels. Levels are corrected for GAPDH. In **b–d**, results represent the averages of three independent experiments.

response to AFX and p27^{kip1}-deficient cells should not. To examine this, we infected immortalized mouse embryo fibroblasts (MEFs) with a retrovirus encoding AFX. In wild-type MEFs, this resulted in induction of p27^{kip1} expression (Fig. 3a), inhibition of cyclin-E-associated kinase activity (Fig. 3a), strong inhibition of BrdU incorporation (Fig. 3b) and increased binding of p27^{kip1} in cyclin E complexes (Fig. 3a). When the same retrovirus was used to infect immortalized pRb-deficient MEFs, a similar inhibition of BrdU incorporation was seen (Fig. 3b). More importantly, in immortalized p27^{kip1}-deficient MEFs, inhibition of BrdU incorporation (Fig. 3b) and cyclin-E-associated kinase activity (Fig. 3d) by AFX were greatly reduced, although inhibition of BrdU incorporation by p16^{ink4a} occurred normally in these cells. BrdU incorporation and cyclin-E-associated kinase activity were also differentially affected in primary MEFs established from wild-type and p27 null mice (Fig. 3c), ruling out the possibility that genetic variation occurring during immortalization gave rise to the observed differences. In addition, inhibition of endogenous PI(3)K/PKB signalling by addition of LY294002 caused a very efficient inhibition of cyclin-E-associated kinase activity in wild-type MEFs. This inhibitory activity was severely reduced in the p27^{kip1}-deficient MEFs, very similar to our observations after ectopic expression of AFX (Fig. 3d). Moreover, LY294002-mediated inhibition of cyclin E/cdk2 activity requires endogenous AFX-like activity, as expression of the AFX.DB interfering mutant reduced LY294002-induced inhibition of cyclin E/cdk2 complexes in A14 cells (Fig. 3e). Thus, our data show that the inhibition of cyclin E/cdk2 complexes by AFX and the endogenous PI(3)K/PKB signalling pathway both act mainly through p27^{kip1}. Some residual inhibition of cyclin-E-associated kinase activity still appears in response to AFX in p27-deficient MEFs, and therefore other cell-cycle targets of AFX besides p27^{kip1} may exist. Nevertheless, these data clearly show that p27^{kip1} is a major target of the PI(3)K/PKB/AFX signalling pathway in regulation of cellular proliferation.

To examine whether the observed increase in p27^{kip1} protein levels by AFX is due to a stabilization of the protein, we analysed p27^{kip1} degradation in the absence or presence of AFX by western blot, using cycloheximide-treated cells (Fig. 4a) or a ³⁵S-methionine pulse-chase experiment (not shown). The kinetics of p27^{kip1} degradation were similar in both cases ($t_{1/2} \approx 45$ min), although p27^{kip1} protein levels were elevated in AFX-expressing cells (Fig. 4a). To study whether AFX enhances transcription of the p27^{kip1} gene, we transiently transfected Jurkat and A14 cells with a luciferase reporter construct driven by the p27^{kip1} promoter, which contains multiple putative AFX-binding sites^{5,17}, together with the various AFX plasmids. AFX and AFX-A3 increased luciferase activity up to 20-fold (Fig. 4b). Insulin treatment of A14 cells expressing AFX inhibited luciferase activity by 50%, similar to the observed inhibition of insulin on other AFX-regulated promoter constructs⁵⁻⁷. These experiments show that AFX can regulate p27^{kip1} promoter activity. Next, we performed northern blots using RNA isolated from wild-type MEFs infected with the pBabe.AFX or pBabe.puro retroviruses. Cells expressing AFX displayed a moderate, but consistent, increase in p27^{kip1} mRNA levels (Fig. 4c). In addition, insulin treatment of A14 cells results in the phosphorylation and inactivation of endogenous Forkheads (Fig. 1d), and a consequent reduction in p27^{kip1} mRNA levels within 2 h (Fig. 4d).

The data presented here demonstrate that AFX is important in regulating cellular proliferation. AFX integrates signals from PI(3)K/PKB signalling and Ras/Ral signalling to regulate transcription of p27^{kip1}. Indeed, p27^{-/-} cells are significantly less inhibited by AFX activity than their p27^{+/+} counterparts, showing that p27^{kip1} is a major target of AFX-like Forkhead proteins. Our finding that these Forkhead transcription factors are predominantly involved in the regulation of cell-cycle progression indicates that the proclaimed role of PI(3)K/PKB signalling in tumorigenesis may need refinement. The disruption of the PI(3)K/PKB/Forkhead transcription

factor pathway during tumorigenesis may override a cell-cycle block rather than bestow protection from apoptosis. In this respect, it is interesting to note that the ability to arrest in G₁ is coupled with the ability to undergo apoptosis in certain cells^{18,19}. This indicates that a G₁ arrest may precede apoptosis, similarly to G₁ arrest preceding cell differentiation. □

Methods

Cloning, plasmids and cells

pBabe.AFX was created by ligating a Klenow-blunted *Pst*I/*Kpn*I fragment from pMT2HA-AFX⁵ into an *Eco*RI-cut and Klenow-blunted dephosphorylated pBabe.puro plasmid²⁰. pMT2HA-AFX.A3 was created by mutating T28, S193 and S258 to A using primers as described⁵. pMT2HA-AFX.ΔDB was created by deleting basepairs +313 to +537 using PCR-based mutagenesis with the primer ΔDB (5'-GAAATCAGTCATATGCAGAAGGA GGCAAGAGCGGCAAAGC-3'). pMT2HA-DB was created by ligating a *Kpn*I/*Xba*I fragment from PRP261-AFX.DB⁵ into *Kpn*I/*Xba*I-cut pMT2HA. Myc-tagged pRb and pCMV.p27^{kip1} were a gift from R. Bernards. pBabe-p16^{ink4a}, pCMV.CD20, pCM.cyclinD1, pCMV.cyclinE, pCMV.cdk4, pCMV.cdk2, pCMV.largeT and pCMV.p16^{ink41} have all been described²⁰. pECE.FKHR-L1 and pECE.FKHR-L1.A3 were a gift from M. Greenberg⁶. pAlter-MAX-FKHR and pAlter-MAX-FKHR.A3 were a gift from T. Unterman⁷. p27GL-1609 and was a gift from I. Touw¹⁷. Immortalized wild-type, pRb-deficient and p27^{kip1}-deficient MEFs were a gift from D. Peeper¹³. Early passage primary wild-type and p27^{kip1}-deficient MEFs were a gift from B. Scheien.

Cell culture

A14 (NIH3T3 overexpressing human insulin receptor), U87MG, RR2, U20S, Phoenix and MEF cells were grown in DMEM with standard supplements. Jurkat cells were grown on RPMI-1640 medium. Insulin, LY294002, nocodazole or puromycin were added at a final concentration of 1 μg ml⁻¹, 10 (A14) or 20 μM (U87MG and MEFs), 250 ng ml⁻¹ or 5 μg ml⁻¹, respectively.

Antibodies

The following antibodies were used: 9E10 (Pharmingen), anti-p27^{kip1} (Transduction Laboratories), 12CA5 for HA-tagged proteins, anti-actin (Santa Cruz), and PKB²¹, anti-phosphoS473 PKB (New England Biolabs), anti-phosphoT32 FKHR-L1 (gift from A. Brunet) and an anti-phosphoS193 AFX (a donation from New England Biolabs). Anti-AFX was generated by immunizing rabbits with a C-terminal peptide from AFX conjugated to KLH (Isogen).

Colony formation

Cells were transfected with the expression plasmids in combination with pBabe-puro and pCMV.CD20. Transfection efficiency of individual plates was determined at 48 h after transfection by FACS analysis of CD20 positivity²⁰. Equal numbers of transfected cells were plated and cells were then grown in the presence of 5 μg ml⁻¹ puromycin for two weeks, after which colonies were scored.

Retroviral infection of MEFs

Recombinant retroviruses were obtained by transfection of the relevant retroviral construct into Phoenix cells. Conditioned medium from these cells was collected 36 h after transfection, filtered and diluted 1:1 with fresh medium and 6 μg ml⁻¹ polybrene was added. This mixture was used to infect exponentially growing MEFs by two consecutive rounds of infection. We routinely obtained 80–90% infection efficiency, as determined by puromycin selection.

Cell-cycle distribution

To analyse cell-cycle redistribution in response to ectopic expression of AFX-like proteins, cells were transfected with the appropriate expression plasmids in combination with CMV.CD20. Thirty-six hours after transfection DNA profiles of CD20-positive cells were obtained by flow cytometry²⁰ and analysed using ModFit software (Becton Dickinson). Retrovirally infected MEFs were pulsed with 1 μM BrdU for 30 min at 48 h after infection and analysed by bivariate flow cytometry as described²².

Magnetic activated cell sorting (MACS)

Cells transfected with pCMV.CD20 and the indicated expression plasmids were stained with anti-CD20 monoclonal antibody (DAKO) followed by staining with goat-anti-mouse microbeads (Miltenyi Biotec). Transfected (CD20-positive) and non-transfected (CD20-negative) cells were then separated on miniMACS separation units (Miltenyi Biotec) according to the supplier's protocol. Transfected (elution) and untransfected (flow-through) fractions were obtained from each transfection and cells were lysed in the appropriate buffers. Positively selected cells were >85% CD20-positive, and <10% CD20-positive cells remained in the negatively selected fraction.

Kinase assays

MACS-separated or retrovirally infected cells were used for determination of cdk2- or cyclin-E-associated kinase activity as described²².

Northern blotting

Total RNA was extracted from wild-type MEFs infected with the indicated retroviruses using standard guanidine isothiocyanate extraction method. We electrophoresed 20 µg of RNA and blotted it onto nitrocellulose and probed for p27^{kip1} and GAPDH messenger using radiolabelled cDNA probes.

p27^{kip1} protein stability

MEFs infected with the indicated retroviruses were treated 48 h post-infection with 10 µg ml⁻¹ cycloheximide for 30 min. Cells were collected at 0, 30, 65, 125, 240 and 460 min after cycloheximide treatment, and p27^{kip1} and actin levels were determined by western blotting of total cell lysates.

Luciferase assay

Cells were transfected with 1 µg (A14) or 5 µg (Jurkat) p27GL-1609 plasmid along with the various AFX plasmids and 1 or 5 µg pCMV-lacZ. Luciferase and lacZ measurements were performed as described⁵.

Received 4 January; accepted 23 February 2000.

1. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A C. *elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
2. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. daf-16: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
3. Ogg, S. *et al.* The Forkhead transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
4. Paradis, S. & Ruvkun, G. *Caenorhabditis elegans* Akt/PKB transduces insulin receptor-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. *Genes Dev.* **12**, 2488–2498 (1998).
5. Kops, G. J. P. L. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
6. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
7. Guo, S. *et al.* Phosphorylation of serine 256 by protein kinase B disrupts transactivation by FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence. *J. Biol. Chem.* **274**, 17184–17192 (1999).
8. Rena, G., Guo, S., Cichy, S. C., Unterman, T. G. & Cohen, P. Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. *J. Biol. Chem.* **274**, 17179–17183 (1999).
9. Tang, E. D., Nuñez, G., Barr, F. G. & Guan, K. L. Negative regulation of the forkhead transcription factor FKHR by Akt. *J. Biol. Chem.* **274**, 16741–16746 (1999).
10. Furnari, F. B., Huang, H. J. & Cavenee, W. K. The phosphoinositide phosphatase activity of PTEN mediates a serum-sensitive G1 growth arrest in glioma cells. *Cancer Res.* **58**, 5002–5008 (1998).
11. Li, D. M. & Sun, H. PTEN/MMAC1/TEP1 suppresses the tumorigenicity and induces G1 cell cycle arrest in human glioblastoma cells. *Proc. Natl Acad. Sci. USA* **95**, 15406–15411 (1998).
12. Roche, S., Koegl, M. & Courtneidge, S. A. The phosphatidylinositol 3-kinase alpha is required for DNA synthesis induced by some, but not all, growth factors. *Proc. Natl Acad. Sci. USA* **91**, 9185–9189 (1994).
13. Peeper, D. S. *et al.* Ras signalling linked to the cell-cycle machinery by the retinoblastoma protein. *Nature* **386**, 177–181 (1997).
14. Haas-Kogan, D. *et al.* Protein kinase B (PKB/Akt) activity is elevated in glioblastoma cells due to mutation of the tumor suppressor PTEN/MMAC. *Curr. Biol.* **8**, 1195–1198 (1998).
15. Burgering, B. M. *et al.* Insulin stimulation of gene expression mediated by p21ras activation. *EMBO J.* **10**, 1103–1109 (1991).
16. Ohtsubo, M., Theodoras, A. M., Schumacher, J., Roberts, J. M. & Pagano, M. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol. Cell. Biol.* **15**, 2612–2624 (1995).
17. Kwon, T. K., Nagel, J. E., Buchholz, M. A. & Nordin, A. A. Characterization of the murine cyclin-dependent kinase inhibitor gene p27^{kip1}. *Gene* **180**, 113–120 (1996).
18. Lissy, N. A. *et al.* TCR antigen-induced cell death occurs from late G1 phase cell cycle check point. *Immunity* **8**, 57–65 (1998).
19. Boonen, G. J. J. C. *et al.* Cyclin D3 regulates proliferation and apoptosis of leukemic T cell lines. *J. Biol. Chem.* **274**, 34676–34682 (1999).
20. Medema, R. H., Herrera, R. E., Lam, F. & Weinberg, R. A. Growth suppression by p16ink4 requires functional retinoblastoma protein. *Proc. Natl Acad. Sci. USA* **92**, 6289–6293 (1995).
21. Burgering, B. M. T. & Coffey, P. J. Protein kinase B (c-Akt) in phosphatidylinositol 3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
22. Medema, R. H., Klompmaker, R., Smits, V. A. J. & Rijkse, G. p21waf1 can block cells at two points in the cell cycle, but does not interfere with processive DNA-replication or stress-activated kinases. *Oncogene* **16**, 431–441 (1998).

Acknowledgements

We thank B. van Oirschot, R. Klompmaker, V. Smits and L. de Vries-Smits for technical assistance, M. Greenberg, A. Brunet, T. Unterman, R. Bernards, I. Touw, D. Peeper and B. Scheien for providing useful reagents, and K. Reedquist for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to B.M.T.B. (e-mail: b.m.t.burgering@med.uu.nl).

Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage

Takeshi Nishikawa^{*,†}, Diane Edelstein^{*}, Xue Liang Du^{*}, Sho-ichi Yamagishi^{*}, Takeshi Matsumura^{*}, Yasufumi Kaneda[‡], Mark A. Yorek[§], David Beebe^{||}, Peter J. Oates^{||}, Hans-Peter Hammes[¶], Ida Giardino^{*} & Michael Brownlee^{*}

^{*} Albert Einstein College of Medicine, Diabetes Research Centre, 1,300 Morris Park Ave, Bronx, New York 10461, USA

[‡] Division of Gene Therapy Science, Osaka University Medical School, Suita Osaka, 5650871 Japan

[§] Department of Internal Medicine, University of Iowa, Iowa City, Iowa 52246, USA

^{||} Department of Metabolic Diseases Central Research Division, Pfizer Inc., Groton, Connecticut 06340, USA

[¶] Justus-Liebig University III Medical Department, Giessen, D-35385, Germany

Diabetic hyperglycaemia causes a variety of pathological changes in small vessels, arteries and peripheral nerves¹. Vascular endothelial cells are an important target of hyperglycaemic damage, but the mechanisms underlying this damage are not fully understood. Three seemingly independent biochemical pathways are involved in the pathogenesis: glucose-induced activation of protein kinase C isoforms²; increased formation of glucose-derived advanced glycation end-products³; and increased glucose flux through the aldose reductase pathway⁴. The relevance of each of these pathways is supported by animal studies in which pathway-specific inhibitors prevent various hyperglycaemia-induced abnormalities^{3,5–7}. Hyperglycaemia increases the production of reactive oxygen species inside cultured bovine aortic endothelial cells⁸. Here we show that this increase in reactive oxygen species is prevented by an inhibitor of electron transport chain complex II, by an uncoupler of oxidative phosphorylation, by uncoupling protein-1 and by manganese superoxide dismutase. Normalizing levels of mitochondrial reactive oxygen species with each of these agents prevents glucose-induced activation of protein kinase C, formation of advanced glycation end-products, sorbitol accumulation and NFκB activation.

Intracellular glucose oxidation begins with glycolysis in the cytoplasm, which generates NADH and pyruvate. Cytoplasmic NADH can donate reducing equivalents to the mitochondrial electron transport chain through two shuttle systems, or it can reduce pyruvate to lactate, which leaves the cell to provide a substrate for hepatic gluconeogenesis. Pyruvate can also be transported into the mitochondria, where it is oxidized by the tricarboxylic acid (TCA) cycle to produce carbon dioxide, water, four molecules of NADH and one molecule of FADH₂. Mitochondrial NADH and FADH₂ provide energy for ATP production through oxidative phosphorylation by the electron transport chain.

Electron flow through the mitochondrial electron transport chain is carried out by four inner-membrane-associated enzyme complexes, with cytochrome c and the mobile carrier ubiquinone⁹. NADH derived from both cytosolic glucose oxidation and mitochondrial TCA cycle activity donates electrons to NADH:ubiquinone oxidoreductase (complex I). Complex I ultimately transfers its electrons to ubiquinone.

Ubiquinone can also be reduced by electrons donated from several FADH₂-containing dehydrogenases, including succinate:ubiquinone oxidoreductase (complex II) and glycerol-3-phosphate

[†]Present address: Kumamoto University School of Medicine, Department of Medicine, Kumamoto, Japan.

dehydrogenase. Electrons from reduced ubiquinone are then transferred to ubiquinol:cytochrome c oxidoreductase (complex III) by the ubisemiquinone radical-generating Q cycle¹⁰. Electron transport then proceeds through cytochrome c, cytochrome c oxidase (complex IV) and, finally, molecular oxygen.

Electron transfer through complexes I, III and IV generates a proton gradient that drives ATP synthase (complex V). When the electrochemical potential difference generated by this proton gradient is high, the life of superoxide-generating electron transport intermediates such as ubisemiquinone is prolonged. There appears to be a threshold value above which superoxide production is markedly increased¹¹. There are two main sites of superoxide generation in the inner mitochondrial membrane: NADH dehydrogenase at complex I, and the interface between ubiquinone and complex III^{9,12}.

Cytosolic NADH is transferred into mitochondria predominantly through the malate–aspartate shuttle, whereas mitochondrial NADH and FADH₂ are generated by the TCA cycle from cytosolic pyruvate. To determine the role of the malate–aspartate shuttle in hyperglycaemia-induced production of reactive oxygen species (ROS), we incubated bovine aortic endothelial cells with amino-oxyacetate, a well characterized inhibitor of the malate–aspartate shuttle. No effect on hyperglycaemia-induced ROS production was observed (Fig. 1). In contrast, inhibition of glycolysis-derived pyruvate transport into mitochondria by 4-hydroxycyanocinnamic acid completely inhibited hyperglycaemia-induced ROS production (Fig. 1). Flux through the TCA cycle, measured by ¹⁴CO₂ production from [U-¹⁴C]glucose (ref. 13), was increased 2.2-fold by 30 mM glucose (0.183 ± 0.005 nmol mg⁻¹ min⁻¹ versus 0.084 ± 0.005 nmol mg⁻¹ min⁻¹ for 5 mM glucose). These data indicate that the TCA cycle is the source of increased ROS-generating substrate induced by hyperglycaemia.

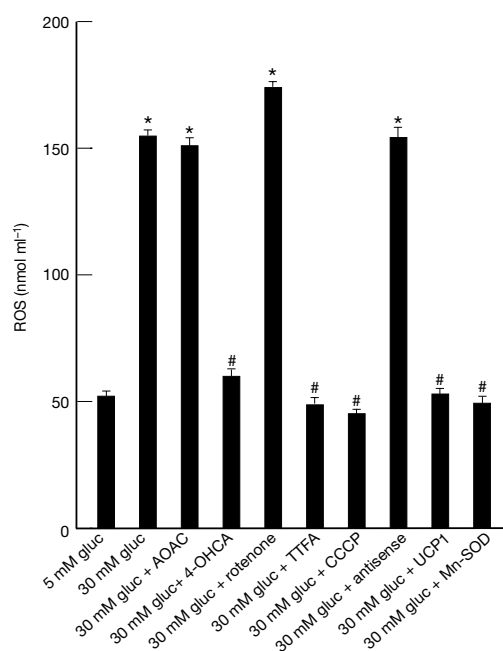


Figure 1 Effect of agents that alter mitochondrial metabolism on formation of hyperglycaemia-induced reactive oxygen species (ROS) in bovine aortic endothelial cells. Cells were incubated ($n = 24$) in 5 mM glucose, 30 mM glucose alone and 30 mM glucose plus either amino-oxyacetic acid (AOAC), α -cyano-4-hydroxycinnamic acid (4-OHCA), rotenone, thenoyltrifluoroacetone (TTFA), carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), antisense, uncoupling protein-1 (UCP1) or manganese superoxide dismutase (Mn-SOD) HVJ-liposomes, and ROS were quantified. Asterisk, $P < 0.01$ compared to cells incubated in 5 mM glucose. Hash, $P < 0.01$ compared to cells incubated in 30 mM glucose.

To determine the site of hyperglycaemia-induced intracellular ROS production, bovine aortic endothelial cells were first incubated with either rotenone, an inhibitor of complex I, thenoyltrifluoroacetone (TTFA), an inhibitor of complex II, or carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), an uncoupler of oxidative phosphorylation that abolishes the mitochondrial membrane proton gradient. Compared with baseline conditions (5 mM glucose), incubation with 30 mM glucose increased ROS production from 52.08 ± 1.03 (5 mM glucose) to 154 ± 1.38 nmol ml⁻¹ (Fig. 1). Rotenone did not reduce this increased ROS production, whereas both TTFA and CCCP completely prevented the effect of hyperglycaemia (Fig. 1). Neither TTFA nor CCCP reduced the increased rate of glucose flux through glycolysis (5.47 ± 0.60 nmol mg⁻¹ min⁻¹ for 30 mM versus 2.39 ± 0.16 nmol mg⁻¹ min⁻¹ for 5 mM glucose).

As pharmacological inhibitors are not absolutely specific to one cellular target, we also performed experiments using molecular techniques. Overexpression of uncoupling protein-1 (UCP1), a specific protein uncoupler of oxidative phosphorylation capable of collapsing the proton electrochemical gradient¹⁴, also prevented

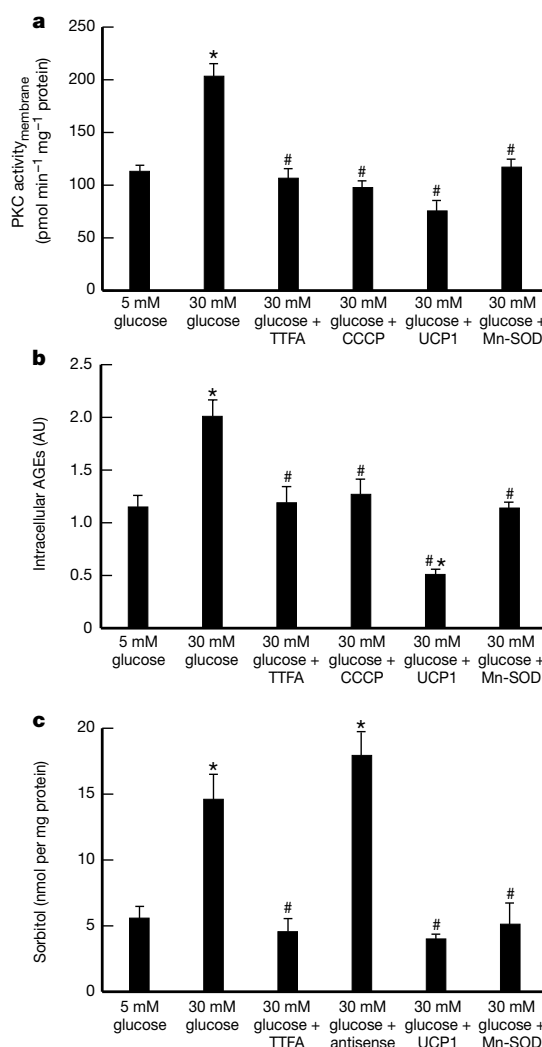


Figure 2 Effect of agents that alter mitochondrial metabolism on the three main pathways of pathogenesis. **a**, Hyperglycaemia-induced PKC activation; **b**, intracellular AGE formation; **c**, sorbitol accumulation. Cells were incubated in 5 mM glucose, 30 mM glucose alone and 30 mM glucose plus either thenoyltrifluoroacetone (TTFA), carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), uncoupling protein-1 (UCP1) or manganese superoxide dismutase (Mn-SOD) HVJ-liposomes, as indicated. Asterisk, $P < 0.01$ compared to cells incubated in 5 mM glucose. Hash, $P < 0.01$ compared to cells incubated in 30 mM glucose.

the effect of hyperglycaemia (Fig. 1). Antisense complementary DNA in the same gene transfer vector did not (Fig. 1). These results show that hyperglycaemia-induced intracellular ROS are produced by the proton electrochemical gradient generated by the mitochondrial electron transport chain. Overexpression of manganese superoxide dismutase (Mn-SOD), the mitochondrial form of this antioxidant enzyme¹⁵, also prevented the effect of hyperglycaemia (Fig. 1). This result shows that superoxide is the reactive oxygen radical produced by this mechanism.

Activation of protein kinase C (PKC) appears to be important in the pathogenesis of diabetic complications². Therefore, we tested the effects of TTFA, CCCP, UCP1 and manganese superoxide dismutase on hyperglycaemia-induced activation of PKC (Fig. 2a). Each completely inhibited PKC activation, indicating that mitochondrial superoxide overproduction may initiate the hyperglycaemia-induced *de novo* synthesis of diacylglycerol or phosphatidylcholine hydrolysis that activates PKC².

Excessive production of advanced glycation end-products (AGEs) also appears to be important in the pathogenesis of diabetic complications³. In bovine aortic endothelial cells, hyperglycaemia increases intracellular AGEs primarily, if not exclusively, by increasing the concentration of AGE-forming methylglyoxal¹⁶. We therefore investigated the effects of TTFA, CCCP, UCP1 and Mn-SOD on hyperglycaemia-induced formation of intracellular methylglyoxal-derived AGEs. Each of these agents completely prevented hyperglycaemia-induced formation of intracellular AGEs (Fig. 2b), indicating that mitochondrial superoxide initiates intracellular AGE formation. As methylglyoxal is formed by fragmentation of glyceraldehyde-3-phosphate, this dependency on increased mitochondrial superoxide production may reflect the well described reversible inhibition of glyceraldehyde-3-phosphate dehydrogenase by ROS, which increases glyceraldehyde-3-phosphate levels¹⁷.

Increased flux through the polyol pathway is a third mechanism that appears to be important in the pathogenesis of diabetic complications⁴. In this pathway, elevated glucose concentration

results in increased production of sorbitol by the enzyme aldose reductase. Sorbitol levels were 2.6-fold higher than baseline (5 mM glucose) when endothelial cells were incubated in 30 mM glucose (Fig. 2c). Sorbitol accumulation at both glucose concentrations was completely inhibited by the specific aldose reductase inhibitor zopolrestat, indicating that, in bovine aortic endothelial cells, sorbitol is derived exclusively from aldose reductase activity (data not shown). Hyperglycaemia-induced sorbitol accumulation was completely prevented by TTFA, UCP1 and Mn-SOD (Fig. 2c), indicating that mitochondrial superoxide overproduction stimulates aldose reductase activity. This observation is consistent with recent data indicating that aldose reductase activity is reversibly downregulated by nitric oxide modification of a cysteine residue in the enzyme's active site¹⁸, and that ROS appear to reduce nitric oxide levels in diabetic endothelium¹⁹.

Hyperglycaemia also activates NFκB, in part by activation of PKC²⁰. We evaluated the effect of TTFA and CCCP on activation of NFκB using an electrophoretic mobility shift assay. Incubation with 30 mM glucose increased NFκB activation 1.8-fold (Fig. 3A), from 43.97 ± 5.08 to 80.54 ± 6.30 arbitrary units (AU). As previously reported in bovine aortic endothelial cells²⁰, this activation by glucose was specific for the NFκB p50 subunit (lower band of NFκB binding activity, Fig. 3A). NFκB activation by 30 mM glucose was completely inhibited by both TTFA (41.18 ± 2.32 AU) and CCCP (42.68 ± 2.30 AU). We tested the effects of UCP1 and Mn-SOD on NFκB activation using a fluorescent *in situ* DNA-protein binding assay (Fig. 3B). Incubation with 30 mM glucose increased NFκB activation 3.5-fold, from 535 ± 5.6 to 1859 ± 18.4 AU per cell (Fig. 3Ba, b). Antisense had no effect (Fig. 3Bc), whereas both UCP1 (Fig. 3Bd) and Mn-SOD (Fig. 3Be) completely prevented hyperglycaemia-induced NFκB activation (to 513 ± 5.9 and 445 ± 6.5 AU per cell, respectively). These observations indicate that glucose activates endothelial cell NFκB by inducing overproduction of mitochondrial superoxide.

Our results show that a single unifying mechanism of induction, increased production of superoxide by the mitochondrial electron transport chain, is a causal link between elevated glucose and each of the three main pathways responsible for hyperglycaemic damage. The data reported here provide the basis for development of new pharmacological agents that either inhibit hyperglycaemia-induced mitochondrial superoxide overproduction or dismutate the superoxide so produced. Such agents may help to prevent the development and progression of diabetic complications. □

Methods

Cell-culture conditions

Confluent bovine aortic endothelial cells (passage 4–10) were maintained in Eagle's minimal essential medium (MEM, Gibco) containing 0.4% fetal bovine serum (FBS), essential and nonessential amino acids and antibiotics. Cells were incubated with either 5 mM glucose, 30 mM glucose, 30 mM glucose plus 100 μM aminooxyacetic acid (AOAC), 250 μM α-cyano-4-hydroxycinnamic acid (4-OHCA), 5 μM rotenone, 10 μM thenoyl-trifluoroacetone (TTFA), 0.5 μM carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), antisense, uncoupling protein-1 (UCP1) or manganese superoxide dismutase (Mn-SOD) HVJ-liposomes, as indicated. Inhibitors were purchased from Sigma. Five microlitres of each HVJ-liposome were added for experiments in 96-well plates (80,000 cells), 10 μl for experiments in 24-well plates (150,000 cells), 100 μl for experiments in 60-mm plates (2×10^6 cells), and 300 μl for experiments in 100-mm plates (6×10^6 cells). Cells were washed after 2 h incubation. Transfection efficiency was >90% as assessed by fluorescence-activated cell sorting analysis of eGFP expression.

Intracellular reactive oxygen species

The intracellular formation of reactive oxygen species was detected using the fluorescent probe CM-H₂DCFDA (Molecular Probes). Cells (1×10^5 ml⁻¹) were loaded with 10 μM CM-H₂DCFDA, incubated for 45 min at 37 °C, and analysed in an HTS 7000 Bio Assay Fluorescent Plate Reader (Perkin Elmer) using the HTSoft program. ROS production was determined from an H₂O₂ standard curve (10–200 nmol ml⁻¹).

Glucose flux through the TCA cycle

Cells were incubated for 7 days in medium containing 5 mM or 30 mM glucose. 1 μCi ml⁻¹ of U-¹⁴C glucose was added, and the conversion to [¹⁴C]CO₂ was quantified as described¹³.

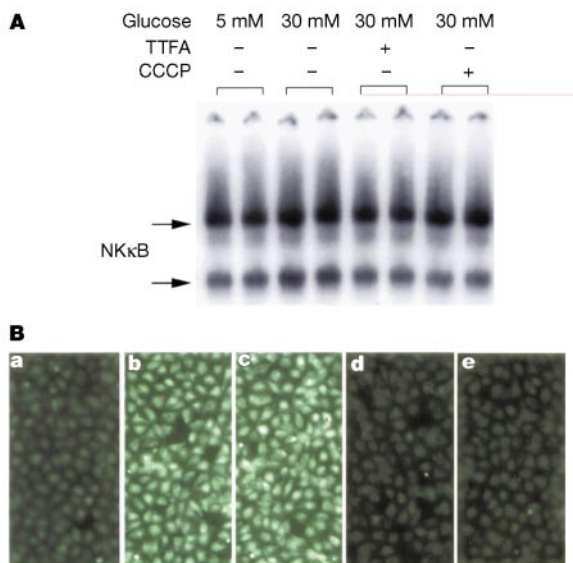


Figure 3 Effect of agents that alter mitochondrial metabolism on hyperglycaemia-induced activation of NFκB. **A**, Cells were incubated in 5 mM glucose, 30 mM glucose alone and 30 mM glucose plus either thenoyltrifluoroacetone (TTFA) or carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), and an electrophoretic mobility shift assay was performed. The lower band of NFκB binding activity is the NFκB p50 subunit. **B**, Cells were incubated in: **a**, 5 mM glucose; **b**, 30 mM glucose; **c**, 30 mM glucose plus antisense; **d**, uncoupling protein-1 (UCP1); or **e**, manganese superoxide dismutase (Mn-SOD) HVJ-liposomes, and a fluorescence *in situ* DNA-protein binding assay was performed.

Glucose flux through glycolysis

Cells were incubated for 7 days in medium containing 5 mM or 30 mM glucose, 30 mM glucose plus TTFA, CCCP, UCP1 or Mn-SOD HVJ liposomes. 10 $\mu\text{Ci ml}^{-1}$ of 5- ^3H glucose was added, and the conversion to [^3H] H_2O was quantified as described¹³.

EBV replicon vectors

Rat UCP1 sense and antisense cDNAs were provided by D. Riquier, Centre National de la Recherche Scientifique-Unité Propre 1511, Meudon, France. Human Mn-SOD cDNA was provided by L. Oberley, University of Iowa College of Medicine. These cDNAs were cloned into the Epstein-Barr virus replicon-based plasmid pEB (ref. 21) and used to prepare HVJ-liposomes.

Preparation of HVJ-liposomes

Cationic HVJ-liposomes were prepared as described²², using 9.75 mg of the dried lipids and 200 μg plasmid DNA. The total volume was adjusted to 1 ml with BSS.

Protein kinase C activity

The assay was performed according to the manufacturer's instructions using the Protein Kinase C Assay System (Life Technologies).

Advanced glycation endproducts

Equal amounts of cell extract protein were used for quantitative immunoblotting performed as described⁸. Methylglyoxal-derived imidazole advanced glycation endproducts were detected using monoclonal antibody 1H7G5 at a 1:10,000 dilution. Immunocomplexes were visualized using an ECF kit according to the manufacturer's instructions (Amersham International) and quantified using a Molecular Dynamics FluorImager and its ImageQuant 4.0 analytical software.

Aldose reductase activity

Cells were incubated as described above for 240 h. Sorbitol content was determined using a temperature-programmed Hewlett-Packard 5890 gas chromatograph interfaced with a model HP3390A integrator as described²³.

NF κ B activation

An electrophoretic mobility shift assay (Gel Shift Assay Kit, Promega) was used according to the manufacturer's instructions. Gels were transferred to Whatman paper, dried and placed in film holders with XAR-5 films at -80°C . Films were scanned using an Ultrascan XL densitometer (Pharmacia). A fluorescence *in situ* DNA-protein binding assay was performed as described²⁴, and fluorescence per cell was determined using IP Lab Spectrum (Scanalytics).

Statistics

Data were analysed using one-factor analysis of variance to compare the means of all the groups. The Tukey-Kramer multiple comparisons procedure was used to determine which pairs of means were different.

Received 24 November 1999; accepted 26 January 2000.

1. National Diabetes Data Group *Diabetes in America* 2nd edn (National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health, USA, 1994).
2. Koya, D. & King, G. L. Protein kinase C activation and the development of diabetic complications. *Diabetes* **47**, 859–866 (1998).

3. Brownlee, M. Advanced protein glycosylation in diabetes and aging. *Annu. Rev. Med.* **46**, 223–234 (1995).
4. Lee, A. Y., Chung, S. K. & Chung, S. S. Demonstration that polyol accumulation is responsible for diabetic cataract by the use of transgenic mice expressing the aldose reductase gene in the lens. *Proc. Natl Acad. Sci. USA* **92**, 2780–2784 (1995).
5. Ishii, H. *et al.* Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor. *Science* **272**, 728–731 (1996).
6. Park, L. *et al.* Suppression of accelerated diabetic atherosclerosis by the soluble receptor for advanced glycation endproducts. *Nature Med.* **4**, 1025–1031 (1998).
7. Sima, A. A., Prashar, A., Zhang, W. X., Chakrabarti, S. & Greene, D. A. Preventive effect of long-term aldose reductase inhibition (ponalrestat) on nerve conduction and sural nerve structure in the spontaneously diabetic Bio-Breeding rat. *J. Clin. Invest.* **85**, 1410–1420 (1990).
8. Giardino, I., Edelstein, D. & Brownlee, M. BCL-2 expression or antioxidants prevent hyperglycemia-induced formation of intracellular advanced glycation endproducts in bovine endothelial cells. *J. Clin. Invest.* **97**, 1422–1428 (1996).
9. Wallace, D. C. Diseases of the mitochondrial DNA. *Annu. Rev. Biochem.* **61**, 1175–1212 (1992).
10. Trumpower, B. L. The protonmotive Q cycle. *J. Biol. Chem.* **265**, 11409–11412 (1990).
11. Korshunov, S. S., Skulachev, V. P. & Starkov, A. A. High protonic potential actuates a mechanism of production of reactive oxygen species in mitochondria. *FEBS Lett.* **416**, 15–18 (1997).
12. Kwong, L. K. & Sohal, R. S. Substrate and site specificity of hydrogen peroxide generation in mouse mitochondria. *Arch. Biochem. Biophys.* **350**, 118–126 (1998).
13. Ishihara, H. *et al.* Effect of Mitochondrial and/or cytosolic glycerol 3-phosphate dehydrogenase overexpression on glucose-stimulated insulin secretion from MIN6 and HIT cells. *Diabetes* **45**, 1238–1244 (1996).
14. Castella, L. *et al.* Stable expression of functional mitochondrial uncoupling protein in Chinese hamster ovary cells. *Proc. Natl Acad. Sci. USA* **87**, 5124–5128 (1990).
15. Manna, S. K. *et al.* Overexpression of manganese superoxide dismutase suppresses tumor necrosis factor-induced apoptosis and activation of nuclear transcription factor- κ B and activated protein-1. *J. Biol. Chem.* **273**, 13245–13254 (1998).
16. Shinohara, M. *et al.* Overexpression of glyoxalase-I in bovine endothelial cells inhibits intracellular advanced glycation product formation and prevents hyperglycemia-induced increases in macro-molecular endocytosis. *J. Clin. Invest.* **101**, 1142–1147 (1998).
17. Knight, R. J., Kofeod, K. F., Schelbert, H. R. & Buxton, D. B. Inhibition of glyceraldehyde-3-phosphate dehydrogenase in post-ischemic myocardium. *Cardiovasc. Res.* **32**, 1016–1023 (1996).
18. Chandra, A., Srivastava, S., Petrash, J. M., Bhatnagar, A. & Srivastava, S. K. Active site modification of aldose reductase by nitric oxide donors. *Biochim. Biophys. Acta* **1341**, 217–222 (1997).
19. Pieper, G. M., Langenstroer, P. & Siebeneich, W. Diabetic-induced endothelial dysfunction in rat aorta: role of hydroxyl radicals. *Cardiovasc. Res.* **34**, 145–156 (1997).
20. Pieper, G. M. & Haq, R. U. Activation of nuclear factor κ B in cultured endothelial cells by increased glucose concentration: prevention by calphostin C. *J. Cardiovasc. Pharmacol.* **30**, 528–532 (1997).
21. Saeki, Y., Wataya-Kaneda, M., Tanaka, K. & Kaneda, Y. Sustained transgene expression *in vitro* and *in vivo* using an Epstein-Barr virus replicon vector system combined with HVJ liposomes. *Gene Therapy* **5**, 1031–1037 (1998).
22. Saeki, Y. *et al.* Development and characterization of cationic liposomes conjugated with HVJ (Sendai virus): reciprocal effect of cationic lipid for *in vitro* and *in vivo* gene transfer. *Human Gene Therapy* **8**, 2133–2144 (1997).
23. Yorek, M. A., Dunlap, J. A. & Lowe, W. L. Osmotic regulation of the Na $^{+}$ /myo-inositol cotransporter and postinduction normalization. *Kidney Int.* **55**, 215–224 (1999).
24. Kurose, I. *et al.* CD18/ICAM-1-dependent oxidative NF κ B activation leading to nitric oxide production in rat kupffer cells cocultured with syngeneic hepatoma cells. *J. Clin. Invest.* **99**, 867–878 (1997).

Acknowledgements

We thank K. Preissner for his critical reading of the manuscript and M. Quon for his gift of plasmid pCIS-eGFP. This research was supported by grants from the National Institutes of Health (USA).

Correspondence and requests for reprints should be addressed to M.B. (e-mail: brownlee@aecom.yu.edu).

Diamonds are for cryo-sectioning

Microscopy and image analysis provide the grist for this week's mill.

C-161 and C-162

Vision Engineering www.visioneng.com

Switchable viewers for Cobra and Lynx instruments

These new viewers promise improved image brilliance for high magnifications without loss of clarity. The 34° angled-view oblique mode can be simply switched to direct mode, and is particularly suitable for applications that benefit from a combination of direct and angled viewing. The oblique mode allows a full 360° view rotation around the centre point, so that critical components or surface features can be inspected in three dimensions without moving the workpiece. The C-161 viewer is finger-tip operated, while the C-162 is motorized. Both can be fitted to existing Lynx or Cobra heads. Used with Vision's Dynascope VS8 scanner, zoom range is $\times 6$ to $\times 40$ in direct, and $\times 4$ to $\times 28$ in oblique view mode.

Reader Service No. 100

NP-1600

Codonics www.codonics.com

Glad to be grey

The NP-1600 can be used with existing colour dye diffusion paper and transparencies, but also offers a greyscale option producing high-quality greyscale prints especially to meet the needs of electron microscopists. The printer's inbuilt software eliminates the need to install printer drivers, and also acts as a web server, allowing access via Netscape Navigator and Internet Explorer. Individual users can control the printer without affecting other users.

Reader Service No. 101

LSM 510 NLO

Carl Zeiss www.zeiss.com/micro

Short-pulse lasers for multi-photon excitation

The LSM 510 NLO version of the LSM 510 confocal laser scanning microscope, used with the Axioskop 2 FS motorized fixed-stage microscope stand, permits simultaneous confocal imaging and electrophysiological measurements. Integrated short-pulse lasers allow multi-photon excitation, protecting living specimens and implementing optimum three-dimensional selectivity and resolution. Image stacks can be recorded over time and functions for display, analysis, three-dimensional reconstruction and archiving are also available. Scanning may be carried out along any curved or freehand line, allowing examination of biological structures such as neuronal and cellular processes. The LSM 510 NLO is suitable for use with all types of GFP. A



The Vision thing: Cobra, on Dynascope V8.

free image browser is available for viewing, sorting, printing and exporting images.

Reader Service No. 102

LaserPix

Bio-Rad Microscience

www.microscopy.bio-rad.com

Confocal image-analysis software

LaserPix software for Bio-Rad confocal image analysis combines full multicolour co-localization analysis with three-dimensional plotting. This facilitates measurement of co-localization coefficients and the display of co-localized pixels and their frequency. An automatic counting and measurement function counts grains or cells and allows 54 different measurement parameters to be displayed, including area, perimeter and density. Data and graphs can be produced from time series images. Total fluorescence intensity from selected areas in a series of images can be compared, providing a useful analysis tool for ploidy or monitoring protein concentration changes within samples.

Reader Service No. 103

cryo-P

Diatome

www.emsdiasum.com

Platform change

The new feature on the latest diamond knife from Diatome is a section pick-up platform, located at the back of the knife, towards the

user. This means that both semithin and ultra-thin sectioning can be done with the same knife. The sections show up clearly on the pick-up platform, made of dark epoxy material, for easy retrieval.

Reader Service No. 104

STEM data sheet

Hitachi Scientific Instruments

www.hsi-europe.com

High-sensitivity elemental analysis

This is the first in a series of technical data sheets from Hitachi for scanning transmission electron microscopes. It covers the application of the HD-2000 STEM to a materials characterization of a platinum-on-carbon catalyst. The data sheet illustrates the use of all three detectors supplied with the HD-2000. Secondary electron and dark- and bright-field transmitted images are used to show dispersion of platinum particles on the carbon substrate. Energy-dispersive X-ray spectra can be acquired from particles as small as 0.5 nm in as little as 10 seconds, thus avoiding the problem of lability in small particles exposed to the electron beam for 30 seconds or more.

Reader Service No. 105

Ergoscan

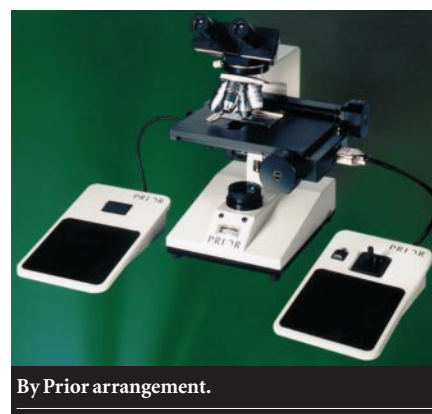
Prior Scientific

www.prior.com

Add joystick control to your microscope

The Ergoscan motorized stage- and focus-control system can be installed on both new and older microscopes. It offers finger-tip control of stage movement using a proportional joystick giving a full range of travel over an area 76×56 mm. The speed of movement is continuously variable from $1 \mu\text{m}$ to 9 mm per second, and a flat roller control provides smooth and precise focusing. Stage and focus controls are available separately.

Reader Service No. 106



By Prior arrangement.



Continuum IR microscope

Nicolet Instruments www.nicolet.co.uk

Versatile microspectroscopy

The Continuum IR microscope is available in a variety of configurations, from a basic microscope to a fully automated system with OMNIC chemical mapping software. The Continuum IR uses infinity-corrected optics for sharp imaging, and differential interference contrast optics to enhance contrast in difficult samples such as biological cells and surface defects. Sample masking allows isolation of specific features. Viewing and data collection can be performed simultaneously, with no need to switch between operational modes. OMNIC software enables automated selection of microsampling parameters and library searching of reference spectra.

Reader Service No. 107

FluoView 300 and 500

Olympus www.olympusamerica.com

Confocal microscope workstations

These fully integrated, multifunctional workstations now offer a greater selection of microscope bases, laser lines and detectors, imaging modes and automated features. Fluorescence signals are detected directly by multiple high-sensitivity 12-bit PMTs. The FluoView 300 features two channels, and FluoView 500 up to four. In each system an additional dedicated PMT provides simultaneous detection of transmitted laser light DIC images. Software includes 3D reconstruction and quantitative image analysis. Suitable for research applications from basic imaging to advanced techniques, the Fluo-View systems can be adapted for use with both upright and inverted Olympus iX70, BX50 and AX70 series.

Reader Service No. 108

LEO 1525

LEO Electron Microscopy www.leo-em.co.uk

Scanning electron microscope

Designed specifically for nano-analysis, the LEO 1525 ultra-high-resolution field emission microscope delivers resolutions of 1.5 nm at 20 kV and 3.5 nm at 1 kV. The design of the specimen chamber allows the common working distance of 8.5 mm to be used for both

ultra-high-resolution imaging and analysis. Navigation is via a high-precision five-axis motorized specimen stage, speed-compensated for magnification and controlled by dual joysticks. The system uses Windows 98 or NT.

Reader Service No. 111

MZ12.5

Leica www.leica-microsystems.com

High performance in stereo

The Leica Microsystems MZ12.5 stereomicroscope offers maximum magnification of $\times 100$, and produces bright, clear images on both light and dark fields. It allows use of a high working distance objective while maintaining wide fields of view. Part of a modular system, the MZ12.5 can easily be customized and upgraded, and is suitable for use with a wide range of Leica accessories.

Reader Service No. 109

MOD-2 S

Halcyonics www.halcyonics.de

Compact vibration isolation

Suitable for AFMs and optical microscopes weighing up to 300 kg, the MOD-2 S offers isolation against all six translational and vibration modes and can be integrated into existing laboratory set-ups. The compact modular design, with two isolation elements measuring only $36 \times 10 \times 11$ cm ($1 \times w \times h$) has no low-frequency resonance and excellent stiffness against forces transferred directly to the application by the hands of an operator or by tubes. Systems suitable for different load and size requirements are available.

Reader Service No. 110

Eclipse E200

Nikon www.nikon.com

Pain-free microscopy

Nikon has applied ergonomic principles to the design of its Eclipse E200 upright micro-



Centurion: go up to $\times 100$ on the Leica MZ12.5.

scope. The focus control and stage handle are positioned equidistant from the operator to allow one-handed operation without twisting the shoulders. The low-profile stage reduces the effort needed to change slides, while the low-inclination-angle eyepiece tube provides a comfortable viewing position. A new re-focusing stage can be lowered by hand to exchange specimens or oil the slide, returning to its original position as soon as the hand is removed. Upper focus limits can be set on the stage movement to prevent damage to the objective and specimen. Suggested applications include epi-fluorescence, phase contrast, simple polarizing methods, photomicrography and face-to-face or side-by-side teaching.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound elsewhere in the journal.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

Peptides
\$14/residue
10 residue minimum

Polyclonal Antibodies - \$1,195

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Look no further
for the best value in
Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS (+44) (0) 1223 839000 • Fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 4

READER ENQUIRY NO. 33

ROLE REVERSAL

Industry may be where the money is, but returning to academia also has a lot to offer

p793

BUILDING BRIDGES

Georgia sees partnerships between business and research as the route to a successful future

p794

INDUSTRIAL GURU

From university to industry and back again — a journey that provides a unique perspective

p794

Universities encourage industrialists to come back to their roots

Returning to the academic world after a spell in industry offers some unique opportunities. Brendan Horton explores the options.

As interaction with industry becomes increasingly important for research universities, academics with industrial experience are finding themselves in demand. This unique combination of skills can help to bridge the cultural gap between the industrial workplace and the laboratory.

Although there are opportunities for researchers to move to academia from industry, having the academic credentials before heading to industry makes an eventual return to the academic world easier, says Dale Compton, Lillian M. Gilbreth distinguished professor of industrial engineering at Purdue University, West Lafayette, Indiana. This is because the measures of scientific achievement are still based almost entirely on publishing and securing grants. "If you spend all of your time in industry, particularly in an environment where you cannot publish, then it's very hard to establish those credentials," says Compton. He spent ten years on the physics faculty at the University of Illinois before undertaking a 16-year stint at the Ford Motor Company, with 14 years as vice-president of research.

An applied perspective

For Compton, the main reason for returning to academia was the chance to interact more closely with young people. He currently teaches the young management-



Williams: aims to teach students about industry.

orientated course in the college of engineering, and his experience in industrial management enables him to provide students with an uncommon perspective on their professional choices.

By contrast, Jim Williams, the Honda professor of materials at the Ohio State University (OSU) in Columbus, says his main objective in returning to academia was to make contributions that he could identify as important "rather than [just]



Rich rewards: university research stands to gain from closer ties with commercial partners.

doing those things that your job requires to be done". For Williams, the best job in a university is that of a full professor. "You have a lot of freedom to do things that you feel are important ... as well as having time to answer the mail, so to speak," he says.

Williams has used his diverse experience from both the public and private sector to his advantage, and it allowed him to end his career on his own terms. Until March 1999, he was the general manager of the materials and process engineering department at General Electric (GE) Aircraft Engines. Before joining GE in 1988, he spent 13 years at Carnegie Mellon University, Pittsburgh, Pennsylvania, as a professor and dean of engineering.

When he left Carnegie Mellon to go to GE, he was attracted by the prospect of being immersed in a practical, product-making business. Now, he hopes that his industrial experience will allow him to change the learning and working environment at OSU, especially in the teaching of undergraduates. Although he feels that many of his academic colleagues might be more knowledgeable than he is in many respects, few have experienced the engineering workplace. Williams believes that he can balance this theoretical base of academic engineering with the

practical knowledge he gained at Boeing, Rockwell and GE.

Williams has found a good deal of support at OSU for someone with his background. "I think that there is broad recognition in the university now that the federal funding picture is not as rosy as it was ten or fifteen years ago," he says. Williams hopes that people who understand how both industry and academia work can help to form better partnerships between the two spheres.

Taking a flexible approach

Considering the consensus that must be arrived at in appointing an endowed chair, flexibility on the part of the institution can be important in closing the deal once a candidate is selected. A willingness to move the starting date of employment, for example, was key to Williams's acceptance of the Honda endowed chair. Without the additional time, Williams could have missed out on collecting important financial incentives available to senior executives at GE who stay until they are 60.

When Williams was considering the chair at OSU, he was still 12 months shy of his sixtieth birthday. However, the university committee offered to defer the appointment,

TERRY ALLEN

▶ allowing Williams to collect on the GE options. This helped to finance his move to OSU, where the academic salary is about a third of the salary and incentive plan he received at GE.

With his stock options secured, Williams became a full professor again at the age of 60½. "If you are leaving industry and you are concerned about compensation, then you probably shouldn't be doing it," he says. ■

Georgia provost Karen Holbrook works closely with Baile and is a strong advocate of this kind of activity, he says. In a previous position as vice-president of research at the University of Florida, Holbrook helped build an incubator facility for start-ups. Stice has found that faculty is also receptive to this kind of approach, and with the president, provost and dean driving it forwards, he hopes that it will be used as a model for future recruitment of eminent scholars to endowed chairs.

This system works, says Stice, because it is structured to optimize the success of both the university and the company. For example, "I'm working on new ways of cloning animals, and I don't have to think about whether I'm working for the company or the university today", he says. All of the intellectual property goes to the university, and the company that is sponsoring the research licenses it from the university.

This is also an exceptional environment for students, most of whom will end up in industry, says Stice. Consequently, it is important for them to get an early taste of what it is like to work in an industrial setting. In this case, that company might be located in the next lab or down the hall. "Obviously, industry is not right for everyone," he says, but in this environment students are able to make more informed career choices. ■

*Brendan Horton is a freelance science journalist.
e-mail: b.horton@naturedc.com*

Georgia realizes the commercial potential of science

Science in the state of Georgia is undergoing a dramatic transformation as stronger links are forged between business and academia. The organization behind this work is the Georgia Research Alliance (GRA), which has brought the state's government, business community and research universities into a partnership that enhances scientific development and subsequent commercialization. "In Georgia, we believe that our research universities are powerful engines for economic development," says GRA president Michael Cassidy.

The GRA offers resources to foster, and perhaps accelerate, Georgia's economic development through activities such as helping to finance new university buildings. For example, Clifton Baile, a GRA eminent scholar in agricultural biotechnology at the University of Georgia, has been involved in a \$28 million biotechnology initiative. This covers five different sites, the largest of which is a \$15 million building at the University of Georgia. This new facility has areas for scientific discovery and product commercialization, and a third of the building is available for lease to start-up companies.

The GRA technology partnership programme gives support to universities that match commercial money for start-up activities. Funding from the GRA is also available for developing core laboratory facilities, such as a \$16 million, three-year proposal for laboratories in genomics and proteomics. In addition, the GRA supplies loaned equipment to start-ups within university incubators.

Wanted: eminent scholars

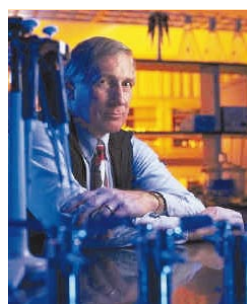
The GRA also provides funding for eminent scholars and endowed chairs in the state's universities. This is an especially effective contribution, says Baile, as it is used exclusively to attract talent and technology from outside Georgia. One such position is now being advertised in the area of animal genomics, and specifically designed facilities are part of the package.

Baile, who worked for 13 years in industrial research at Monsanto (see box), says that the eminent scholar programme was one of the mechanisms used to attract Steve Stice to the University of Georgia in 1998. Baile recruited him from Advanced Cell Technolo-

gies, a company that Stice co-founded while an adjunct professor at the University of Massachusetts in Amherst.

In addition to accepting the \$1.5 million endowed chair, Stice is employed by ProLinia, one of the start-up companies supported by the university and the GRA. This unusual dual appointment is structured so that Stice spends 51 per cent of his time as a professor, with the rest spent at ProLinia. While in Massachusetts, Stice and his senior professor James Robl pioneered the development of cloned transgenic animals. ProLinia was formed to commercialize the use of cloning and transgenics for food-producing animals.

According to Baile, the backing of upper management at the university has made a great difference to the overall success of the GRA initiative. For example, University of



The importance of experience

Few researchers have spent as much time in industry as Clifton Baile and still have enough academic accomplishments to re-enter academia as a professor. Baile may have established his academic credentials before entering industry by working at Harvard University and the University of Pennsylvania on the control of energy balance in animals but, unusually, he was able to continue this work while a director of research at Monsanto. At Monsanto, Baile was one of 130 fellows among the company's 5,000 scientists. "In the fellows programme, I had a responsibility, or a privilege, to continue my scientific interests," he says.

After 13 years of directing research programmes for up to 375 people at Monsanto — during which time he oversaw the development of the controversial drug BST (bovine somatotropin) — Baile was ready to do something else. "That was not just science, and it wasn't just drug development, it was a political war," he says. When the opportunity arose, he took a retirement package from Monsanto that gave him years of financial security over and above what he had anticipated.

When he moved to the University of Georgia he was attracted by the chance to work in the vice-president's office for research as a full professor. Baile thought this set up would be better than being allied to a particular department as it would allow him to be involved in the overall development of research at the university.

"What turned out to be really attractive, and should have been the main reason for coming to Georgia, was the GRA, but I didn't really understand it when I made my decision," says Baile. He expected to be involved with technology transfer, owing to his considerable experience in the commercial sector, but he did not realize that the GRA would provide so much supporting infrastructure. According to Baile, the GRA makes the University of Georgia unusual, if not unique, in what it can offer someone like him.

As a result, Baile now has a very different relationship with the university. "I am a tenured professor, but I'm probably the only tenured professor in the state who is non-salaried," he says. He joined the university without a requirement for salary because he wanted the freedom to use the infrastructure to develop university-commercial relationships and businesses. He has students who are part of his contract with the university and he is responsible for a floor of the animal-science building and its equipment. This floor also houses several of the companies with which Baile works closely. It is unusual and beneficial for students to have exposure to advisers and mentors who are closely tied to the commercialization of scientific discovery, Baile says. ■